



# Stochastic Approximation Finite Element method: analytical formulas for multidimensional diffusion process

Romain Bompis, Emmanuel Gobet

## ► To cite this version:

Romain Bompis, Emmanuel Gobet. Stochastic Approximation Finite Element method: analytical formulas for multidimensional diffusion process. 2013. hal-00842608

**HAL Id: hal-00842608**

**<https://hal.science/hal-00842608>**

Preprint submitted on 9 Jul 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# STOCHASTIC APPROXIMATION FINITE ELEMENT METHOD: ANALYTICAL FORMULAS FOR MULTIDIMENSIONAL DIFFUSION PROCESS

R. BOMPIS\* AND E. GOBET†

**Abstract.** We derive an analytical weak approximation of a multidimensional diffusion process as coefficients or time are small. Our methodology combines the use of Gaussian proxys to approximate the law of the diffusion and a Finite Element interpolation of the terminal function applied to the diffusion. We call this method Stochastic Approximation Finite Element (SAFE for short) method. We provide error bounds of our global approximation depending on the diffusion process coefficients, the time horizon and the regularity of the terminal function. Then we give estimates of the computational cost of our algorithm. This shows an improved efficiency compared to Monte-Carlo methods in small and medium dimensions (up to 10), which is confirmed by numerical experiments.

**Key words.** weak approximation, diffusion processes, Malliavin calculus, finite elements

**AMS subject classifications.** 60H35, 60H07, 65M60, 65Cxx

## 1. Introduction.

**Motivation and contribution of the paper.** We consider for  $d \geq 1$  a  $d$ -dimensional stochastic differential equation (SDE) defined by:

$$X_t = x_0 + \sum_{j=1}^q \int_0^t \sigma_j(s, X_s) dW_s^j + \int_0^t b(s, X_s) ds,$$

where  $(W_t)_{t \geq 0}$  is a standard Brownian motion in  $\mathbb{R}^q$  on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$  with the usual assumptions on the filtration  $(\mathcal{F}_t)_{t \geq 0}$ . Here,  $\sigma$  is a  $d \times q$  matrix and  $b$  is a  $d$ -dimensional vector, their entries being regular and bounded functions. We are interested in deriving analytical approximations of

$$(1.1) \quad \mathbb{E}[h(X_T)]$$

for a given function  $h$ , at least Lipschitz continuous, and a fixed time horizon  $T > 0$ . The explicit calculus of (1.1) is most of the time impossible because the marginal law of the diffusion  $X$  is not known and because of the general form of the function  $h$ . Hence it is usual to perform a numerical method. For low dimension (say  $d \leq 3$ ), we may use PDE schemes since  $(x_0, T) \mapsto \mathbb{E}[h(X_T)]$  solves a linear parabolic PDE but the complexity is increasing very quickly with the dimension  $d$ . For higher dimension, Monte-Carlo methods are preferred, but although almost insensitive to the dimension, they only evaluate the above expectation for a single  $(x_0, T)$ . The aim of this work is to provide an alternative numerical method, based on analytical approximation, and we highlight an approach suiting well to general functions  $h$  without specific form (under reasonable conditions) and to rather general diffusion models. The quick and efficient approximation of SDE distributions is fundamental, it is widely used as a cornerstone of probabilistic algorithms related to the dynamic programming problems which necessitate the evaluation of many nested conditional expectations (for instance, see [1] for optimal stopping problems and [16] for Backward SDEs).

Our subsequent numerical method (called Stochastic Approximation Finite Element, SAFE for short) relies both on the weak approximation of the marginal law of  $X_T$  and on the approximation of the function  $h$ . Firstly, to approximate the law of  $X$ , we consider the Gaussian

\*CMAP, Ecole Polytechnique, Route de Saclay, 91128 Palaiseau cedex, France. Email: [romain.bompis@polytechnique.edu](mailto:romain.bompis@polytechnique.edu).

†Email: [emmanuel.gobet@polytechnique.edu](mailto:emmanuel.gobet@polytechnique.edu). Corresponding author.

proxy process obtained by freezing at  $x = x_0$  the diffusion coefficients:

$$(1.2) \quad X_t^P = x_0 + \sum_{j=1}^q \int_0^t \sigma_j(s, x_0) dW_s^j + \int_0^t b(s, x_0) ds.$$

Using the Proxy principle of [9], we derive a weak approximation in the form (see Theorem 2.1):

$$(1.3) \quad \mathbb{E}[h(X_T)] \approx \mathbb{E}[h(X_T^P)] + \sum_{|\alpha| \leq 3} w_{\alpha, T} \partial_{\epsilon_{\alpha_1} \dots \epsilon_{\alpha_{|\alpha|}}}^{|\alpha|} (\mathbb{E}[h(X_T^P + \epsilon)]) \Big|_{\epsilon=0},$$

where  $\alpha \in \{1, \dots, d\}^{|\alpha|}$  is a multi-index,  $w_{\alpha, T}$  are weights depending explicitly on the SDE coefficients and where the sensitivities  $\partial_{\epsilon_{\alpha_1} \dots \epsilon_{\alpha_{|\alpha|}}}^{|\alpha|} (\mathbb{E}[h(X_T^P + \epsilon)]) \Big|_{\epsilon=0}$  are well defined as soon as the law of  $X_T^P$  is non degenerate. Apart from few specific cases of functions  $h$  (for example if  $h$  has separable variables combined with the independence of the  $X_T^P$  components), the representation (1.3) can not be directly computed in closed forms: however, it can be rewritten in a simple expectation form suitable for simple and direct Monte-Carlo simulations (see Theorem 2.3). To obtain fully analytical formulas, another ingredient is needed. The second step is to approximate the function  $h$  by a local interpolation based on suitable shape functions of Finite Element Methods (see Theorems 2.4-2.5-2.8). Denoting by  $\hat{h}$  the resulting interpolation of  $h$ , the final structure of approximation becomes

$$\mathbb{E}[h(X_T)] \approx \mathbb{E}[\hat{h}(X_T^P)] + \sum_{|\alpha| \leq 3} w_{\alpha, T} \partial_{\epsilon_{\alpha_1} \dots \epsilon_{\alpha_{|\alpha|}}}^{|\alpha|} (\mathbb{E}[\hat{h}(X_T^P + \epsilon)]) \Big|_{\epsilon=0},$$

which accuracy and complexity are given in Theorems 2.6-2.8 and Corollaries 2.7-2.9. The convergence holds as  $b, \sigma$  or  $T$  go to 0 in a suitable sense. The key feature in this methodology is that the interpolation procedure is done in such a way that the calculus of the above expectations is fully explicit and reduces to computations involving the c.d.f. of a one-dimensional Gaussian r.v. and its derivatives (see Subsection 2.2). The flexibility and the accuracy of our formulas allow their use as it stands or alternatively it could serve as a control variates tool to improve Monte-Carlo methods.

**Background results.** We briefly describe the main known approaches to approximate the distribution of a SDE. Time discretization schemes are broadly described in [13]: they consist in replacing  $X$  by an approximation  $\hat{X}$  easier to simulate, the evaluation of  $\mathbb{E}(h(\hat{X}_T))$  is then made using Monte-Carlo simulations. The balance between discretization and integration errors is described in [6].

Alternatively, the cubature on Wiener space by Kusuoka-Lyons-Victoir [15, 18] is a well-established theory. It is based on a smart discrete approximation of the Wiener measure, which leads to solving ODEs in order to approximate  $X$ . The splitting method by Ninomiya-Victoir [19] also reduces to solving ODEs. Clearly, these approaches are different from ours. The quantization method [10] is aimed at approximating the distribution of  $X_T$  with a fixed number of points, optimally w.r.t. a  $L_p$ -norm; for applications to stochastic processes, see for instance [1]. This differs from the current work.

The use of asymptotic methods has been much developed during the recent years, mostly in the fields of mathematical finance. As opposed to our setting, the related works deal mainly with one-dimensional processes and specific  $h$ . Mathematical approaches are numerous, see [7, 3, 17] among others. Here, we address the multi-dimensional case with general functions  $h$ , extending much the setting of previous references.

**Organization of the paper.** In the following, we introduce notations and assumptions that are used throughout the paper. We state in Section 2 the main results of the paper:

- We first provide in Theorem 2.1 a second order weak approximation of  $\mathbb{E}[h(X_T)]$  using the Gaussian proxy, the magnitude of the error being estimated w.r.t. the SDE coefficients and the time horizon  $T$ . The previous approximation, involving correction terms as expectation sensitivities, has an interesting representation as a simple expectation, much suitable for direct Monte-Carlo simulations, see Theorem 2.3.
- We then perform a suitable multilinear interpolation of the function  $h$  in Theorem 2.4, the accuracy results being given in Theorem 2.5 according to the regularity of  $h$ . The resulting formulas are fully explicit.
- We finally establish a final approximation combining both the weak expansion and the interpolation of  $h$  in Theorem 2.6, providing tight error estimates as well a complexity analysis (Corollary 2.7).
- Results are extended in Theorem 2.8 and Corollary 2.9 considering multiquadratic finite elements.

The proof of the error estimates of Theorems 2.1 and 2.5 are respectively given in Sections 3 and 4. Numerical experiments illustrating the performance of our algorithm in comparison to Monte-Carlo methods are presented in Section 5. Appendix A is devoted to the explicit derivation of the corrective terms of the weak expansion provided in Theorem 2.1.

#### Notations, definitions and assumptions.

▷ **Linear algebra.** The  $j$ -th column of a matrix  $A$  will be denoted by  $A_j$  (or  $A_{j,t}$  if  $A$  is a time-dependent matrix) and its  $i$ -th row by  $A^i$ .  $A^*$  denotes the transpose of  $A$  and if it is a squared matrix,  $\det(A)$  stands for its determinant.  $I_m$  denotes the  $m$ -dimensional identity matrix,  $\langle \cdot, \cdot \rangle$  the inner product on  $\mathbb{R}^m$  and  $|\cdot|$  is the Euclidean norm on  $\mathbb{R}^m$ .

▷ **Functions.** As usual,  $C^k(O_1, O_2)$  stands for the set of functions  $g : O_1 \mapsto O_2$  that are  $k$ -times continuously differentiable, where  $O_1, O_2$  are some subsets of Euclidean spaces. Let  $p_1, p_2$  be in  $\mathbb{N} \setminus \{0\}$ . For any function  $g = (g_1, \dots, g_{p_2})^* : [0, T] \times \mathbb{R}^{p_1} \rightarrow \mathbb{R}^{p_2}$ , we denote  $|g|_\infty = \sup_{(t,x) \in [0,T] \times \mathbb{R}^{p_1}} |g(t, x)|$ . If  $g$  is sufficiently differentiable w.r.t. the variable  $x$ , its gradient which takes values in  $\mathbb{R}^{p_2} \otimes \mathbb{R}^{p_1}$  is simply denoted  $\nabla g(t, x) = (\partial_{x_1} g(t, x), \dots, \partial_{x_{p_1}} g(t, x))$ ; when  $p_2 = 1$ , we denote its Hessian matrix by  $H(g)(t, x) = (\partial_{x_i, x_j}^2 g(t, x))_{i,j \in \{1, \dots, p_1\}}$ . Furthermore, we often use the short notation  $\partial^\alpha g(t, x)$  for  $\partial_{x_{\alpha_1} \dots x_{\alpha_{|\alpha|}}} g(t, x)$ , i.e. the partial derivative of  $g$  w.r.t. a multi-index  $\alpha$  according to the space variable. We denote by  $\text{Lip}(\mathbb{R}^d, \mathbb{R})$  the space of Lipschitz functions  $h : \mathbb{R}^d \mapsto \mathbb{R}$  satisfying  $C_{\text{Lip}, h} := \sup_{x, y \in \mathbb{R}^d, x \neq y} \frac{|h(x) - h(y)|}{|x - y|} < +\infty$ .

▷ **About the Gaussian proxy.** Whenever unambiguous, we use the notations  $\sigma_t := \sigma(t, x_0)$  and  $b_t := b(t, x_0)$  for any  $t \in [0, T]$  and we denote by  $\Sigma_t := \sigma_t \sigma_t^*$  the  $d$ -dimensional non-negative definite covariance matrix at time  $t$  associated to the Gaussian process  $X^P$  defined in (1.2). We start with an easy result, which notations are used throughout the paper.

PROPERTY 1.1.

1. The distribution of  $X_T^P$  is normal with mean  $m_T^P = x_0 + \int_0^T b_t dt$  and covariance matrix  $\mathcal{V}_T^P = \int_0^T \Sigma_t dt$ .
2. There is a  $d$ -dimensional orthogonal matrix  $\mathcal{U}_V$  such that  $\mathcal{V}_T^P = \mathcal{U}_V \mathcal{D}_T^P \mathcal{U}_V^{-1}$  where  $\mathcal{D}_T^P := \text{diag}(\lambda_1^2 T, \dots, \lambda_d^2 T)$  is a  $d$ -dimensional diagonal matrix containing the eigenvalues of  $\mathcal{V}_T^P$ .

▷ **Assumption  $(\mathcal{H}_{x_0})$  on  $\sigma$  and  $b$ .**

**$(\mathcal{H}_{x_0})$ -i)**  $\sigma$  and  $b$  are bounded measurable functions from  $[0, T] \times \mathbb{R}^d$  to  $\mathbb{R}^{d \times q}$  and  $\mathbb{R}^d$  respectively, they are twice continuously differentiable w.r.t.  $x$ , with uniformly bounded derivatives, and their second derivatives are locally  $\alpha \in (0, 1]$ -Hölder continuous

w.r.t.  $x$ . We set:

$$\mathcal{M}_1(\sigma, b) = \sum_{\alpha: 1 \leq |\alpha| \leq 2} (|\partial^\alpha \sigma|_\infty + |\partial^\alpha b|_\infty) \quad \text{and} \quad \mathcal{M}_0(\sigma, b) = \max(|\sigma|_\infty, |b|_\infty, \mathcal{M}_1(\sigma, b)).$$

To avoid uninteresting situations, we assume  $\mathcal{M}_0(\sigma, b) > 0$ .

( $\mathcal{H}_{x_0}$ )-ii) There is a constant  $\bar{C}_V \geq 1$  such that

$$\bar{C}_V \mathcal{M}_0(\sigma, b) \geq \max_{i \in \{1, \dots, d\}} \lambda_i \geq \min_{i \in \{1, \dots, d\}} \lambda_i \geq (\bar{C}_V)^{-1} \mathcal{M}_0(\sigma, b).$$

In particular, the matrix  $\mathcal{V}_T^P$  is positive definite.

From ( $\mathcal{H}_{x_0}$ ) and Property 1.1 we easily deduce

PROPERTY 1.2.

1. The distribution of  $X_T^P$  has a density  $f^P(x) = \frac{e^{-\frac{1}{2}(x-m_T^P)^*(\mathcal{V}_T^P)^{-1}(x-m_T^P)}}{(2\pi)^{\frac{d}{2}} \sqrt{\det(\mathcal{V}_T^P)}}$ , such that for any multi-index  $\alpha$

$$(1.4) \quad |\partial^\alpha f^P(x)| \leq C_{\alpha,d} (\mathcal{M}_0(\sigma, b) \sqrt{T})^{-(d+|\alpha|)} \exp\left(-\frac{|x-m_T^P|^2}{C_{\alpha,d} [\mathcal{M}_0(\sigma, b)]^2 T}\right),$$

for a constant  $C_{\alpha,d} > 0$  that depends in a non-decreasing way on  $T$ ,  $\mathcal{M}_0(\sigma, b)$  and  $\bar{C}_V$ .

2. For any measurable function  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$  exponentially bounded, define  $\bar{\phi}^P : \epsilon \in \mathbb{R}^d \mapsto \bar{\phi}^P(\epsilon) = \mathbb{E}[\phi(X_T^P + \epsilon)]$ . Then  $\bar{\phi}^P$  is of class  $C^\infty$  and all the derivatives  $\partial_{\epsilon_{\alpha_1} \dots \epsilon_{\alpha_{|\alpha|}}}^{|\alpha|} \bar{\phi}^P(0) := \partial_{\epsilon_{\alpha_1} \dots \epsilon_{\alpha_{|\alpha|}}}^{|\alpha|} \bar{\phi}^P(\epsilon)|_{\epsilon=0}$  exist for any multi-index  $\alpha \in \{1, \dots, d\}^{|\alpha|}$ .

▷ **Miscellaneous.** We use the following notations to state our error estimates throughout the paper:

- " $A = O(B)$ " means that  $|A| \leq CB$  where  $C$  stands for a generic constant that is a non-negative non-decreasing function of the parameters  $d$ ,  $T$ ,  $\mathcal{M}_0(\sigma, b)$ ,  $\mathcal{M}_1(\sigma, b)$  and  $\bar{C}_V$ . Unless made explicit, a generic constant may depend on the test function  $h$ .
- Similarly, if  $A$  is non-negative,  $A \leq_c B$  means that  $A \leq CB$  for a generic constant  $C$ .

Lastly, for a r.v.  $Y \in \mathbb{R}^m$  ( $m \geq 1$ ) and for  $p \geq 1$ ,  $\|Y\|_p = (\mathbb{E}|Y|^p)^{\frac{1}{p}}$  stands for its  $L^p$ -norm.

## 2. Main results.

**2.1. Second order weak approximation.** The model proxy has the advantage to have an explicit Gaussian law and the accuracy of the approximations  $\sigma(t, X_t) \approx \sigma(t, x_0)$  and  $b(t, X_t) \approx b(t, x_0)$  can be justified if  $\mathcal{M}_1(\sigma, b)$ ,  $\mathcal{M}_0(\sigma, b)$  and  $T$  are globally small enough (see Lemma 3.1). Nevertheless, we can not reasonably expect  $\mathbb{E}[h(X_T)] \approx \bar{h}^P(0) = \mathbb{E}[h(X_T^P)]$  to be solely accurate enough and we provide correction terms. To derive them, we make an intensive use of the next interpolated process:

(2.1)

$$X_t^\eta = x_0 + \sum_{j=1}^q \int_0^t \sigma_j(s, \eta X_s^\eta + (1-\eta)x_0) dW_s^j + \int_0^t b(s, \eta X_s^\eta + (1-\eta)x_0) ds, \quad \eta \in [0, 1],$$

so that  $X^{\eta=1} = X$  and  $X^{\eta=0} = X^P$ . Under ( $\mathcal{H}_{x_0}$ )-i), almost surely for any  $t$ ,  $\eta \rightarrow X_t^\eta$  is  $C^2([0, 1], \mathbb{R}^d)$ , see [14]. The dynamics of the two first derivatives  $(\dot{X}_t^\eta := \partial_\eta X_t^\eta)_{t \geq 0}$  and  $(\ddot{X}_t^\eta :=$

$\partial_{\eta^2}^2 X_t^\eta)_{t \geq 0}$  are obtained by a straight differentiation of the SDE satisfied by  $X^\eta$ :

$$\begin{aligned}
 \dot{X}_t^\eta &= \sum_{j=1}^q \int_0^t \nabla \sigma_j(s, x_0 + \eta(X_s^\eta - x_0))(X_s^\eta - x_0 + \eta \dot{X}_s^\eta) dW_s^j \\
 &+ \int_0^t \nabla b(s, x_0 + \eta(X_s^\eta - x_0))(X_s^\eta - x_0 + \eta \dot{X}_s^\eta) ds, \\
 (\ddot{X}_t^\eta)^i &= \sum_{j=1}^q \int_0^t [(X_s^\eta - x_0 + \eta \dot{X}_s^\eta)^* H(\sigma_j^i)(s, x_0 + \eta(X_s^\eta - x_0))(X_s^\eta - x_0 + \eta \dot{X}_s^\eta) \\
 &+ \nabla \sigma_j^i(s, x_0 + \eta(X_s^\eta - x_0))(2\dot{X}_s^\eta + \eta \ddot{X}_s^\eta)] dW_s^j \\
 &+ \int_0^t [(X_s^\eta - x_0 + \eta \dot{X}_s^\eta)^* H(b^i)(s, x_0 + \eta(X_s^\eta - x_0))(X_s^\eta - x_0 + \eta \dot{X}_s^\eta) \\
 &+ \nabla b^i(s, x_0 + \eta(X_s^\eta - x_0))(2\dot{X}_s^\eta + \eta \ddot{X}_s^\eta)] ds, \quad \forall i \in \{1, \dots, d\}.
 \end{aligned}
 \tag{2.2}$$

Setting  $\sigma'_{j,t} := \nabla \sigma_j(t, x_0)$ ,  $\Sigma'_{j,t} := \nabla \Sigma_j(t, x_0)$  and  $b'_t := \nabla b(t, x_0)$ ,  $\dot{X} := \dot{X}^{\eta=0}$  is solution of the SDE:

$$\dot{X}_t = \sum_{j=1}^q \int_0^t \sigma'_{j,s} (X_s^P - x_0) dW_s^j + \int_0^t b'_s (X_s^P - x_0) ds.
 \tag{2.4}$$

Then combining Taylor expansions for the interpolated process  $X^\eta$  and the function  $h$  (here assumed to be smooth enough for the sake of brevity), we propose the following weak stochastic approximation:

$$\mathbb{E}[h(X_T)] = \mathbb{E}[h(X_T^P)] + \mathbb{E}[\nabla h(X_T^P) \dot{X}_T] + \text{Error}_{2,h}^{\text{SA}},
 \tag{2.5}$$

where the explicit calculus of the corrective term  $\mathbb{E}[\nabla h(X_T^P) \dot{X}_T]$  is performed in Proposition A.2 whereas the estimate of the error  $\text{Error}_{2,h}^{\text{SA}}$  is postponed to Section 3. This leads to the following Theorem (stated for only Lipschitz function  $h$ ).

**THEOREM 2.1. (Second order weak approximation using the Gaussian proxy).**  
Assume  $(\mathcal{H}_{x_0})$  and suppose that  $h \in \text{Lip}(\mathbb{R}^d, \mathbb{R})$ . Then we have:

$$\mathbb{E}[h(X_T)] = \mathbb{E}[h(X_T^P)] + \text{Cor}_{2,h} + \text{Error}_{2,h}^{\text{SA}},
 \tag{2.6}$$

where:

$$\begin{aligned}
 \text{Cor}_{2,h} &= \nabla \bar{h}^P(0) \int_0^T b'_t \left( \int_0^t b_s ds \right) dt + \sum_{i,j=1}^d \partial_{\epsilon_i, \epsilon_j}^2 \bar{h}^P(0) \left[ \int_0^T (b'_t)^i \left( \int_0^t \Sigma_{j,s} ds \right) dt \right. \\
 &+ \left. \frac{1}{2} \int_0^T (\Sigma_{j,t}^i)' \left( \int_0^t b_s ds \right) dt \right] + \frac{1}{2} \sum_{i,j,k=1}^d \partial_{\epsilon_i, \epsilon_j, \epsilon_k}^3 \bar{h}^P(0) \int_0^T (\Sigma_{j,t}^i)' \left( \int_0^t \Sigma_{k,s} ds \right) dt,
 \end{aligned}
 \tag{2.7}$$

recalling  $b'_t = \nabla b(t, x_0)$  and  $(\Sigma_{j,t}^i)' = \nabla[\sigma \sigma^*]_{j,t}^i(t, x_0)$ . The stochastic approximation error term is estimated as follows:

$$|\text{Error}_{2,h}^{\text{SA}}| \leq_c C_{\text{Lip},h} \mathcal{M}_1(\sigma, b) [\mathcal{M}_0(\sigma, b)]^2 T^{\frac{3}{2}}.
 \tag{2.8}$$

Remind that  $\text{Cor}_{2,h}$  is well defined whatever the smoothness of  $h$  (see Property 1.2).

REMARK 2.2. *The weak approximation is constituted by a leading order  $\bar{h}^P(0)$  plus a sum of weighted sensitivities, i.e. derivatives of  $\bar{h}^P$  at zero, up to the third order. The error is of order 3 w.r.t. the standard deviation  $M_0(\sigma, b)\sqrt{T}$  and is null if  $M_1(\sigma, b) = 0$  or if  $C_{\text{Lip},h} = 0$  (i.e.  $h$  is constant). That justifies the label of second order weak approximation. When  $h(x) = \phi(\sum_{i=1}^d \eta_i x_i)$  with  $\eta_i \geq 0$ , the above expansion coincides with that of [9, Theorem 2.1] related to averaged diffusions.*

Although the density of the Gaussian proxy is known, the approximation formula (2.6) does not reduce to fully explicit calculations, due to the general form of  $h$ . Nevertheless, we can derive another representation as an expectation of  $h(X_T^P)$  modified by an explicit weight: this is easily obtained by transferring the  $\varepsilon$ -differentiation of the expectation  $\bar{h}^P(\varepsilon)$  (associated to correction terms) into a differentiation of the proxy Gaussian density  $f^P$ . This is a somewhat standard argument, in particular regarding the Malliavin calculus applications [20, Section 6.2], we skip details of the derivation. The advantage of this representation as an expectation is to make possible its evaluation by standard Monte-Carlo methods involving only simulations of the Gaussian proxy  $X_T^P$ , see our subsequent numerical experiments.

THEOREM 2.3. *Under the notations and assumptions of Theorem 2.1, the main terms of the stochastic approximation are*

$$(2.9) \quad \mathbb{E}[h(X_T^P)] + \text{Cor}_{2,h} = \mathbb{E}\left[h(X_T^P)\left\{1 + \mathcal{W}[\Sigma, b; x_0]_0^T([\mathcal{V}_0^T]^{-1}(X_T^P - m_T^P))\right\}\right],$$

where we set, for  $Y \in \mathbb{R}^d$ ,

$$\begin{aligned} \mathcal{W}[\Sigma, b; x_0]_0^T(Y) = & \langle Y, \int_0^T b'_t(\int_0^t b_s ds) dt \rangle \\ & + \sum_{i,j=1}^d \left\{ Y^i Y^j - ([\mathcal{V}_0^T]^{-1})_{ij}^i \right\} \left[ \int_0^T (b'_t)^i(\int_0^t \Sigma_{j,s} ds) dt + \frac{1}{2} \int_0^T (\Sigma_{j,t}^i)'(\int_0^t b_s ds) dt \right] \\ & + \frac{1}{2} \sum_{i,j,k=1}^d \left\{ Y^i Y^j Y^k - Y^k ([\mathcal{V}_0^T]^{-1})_{ij}^i - Y^j ([\mathcal{V}_0^T]^{-1})_{ik}^i - Y^i ([\mathcal{V}_0^T]^{-1})_{jk}^i \right\} \int_0^T (\Sigma_{j,t}^i)'(\int_0^t \Sigma_{k,s} ds) dt. \end{aligned}$$

As an alternative to a Monte-Carlo evaluation based on (2.9), we provide in the following Subsection a new numerical method to approximate the expansion formula (2.6) taking advantage of a multilinear interpolation with hat functions, which theoretical accuracy is given according to the  $h$ -smoothness. The extension to multiquadratic interpolation is presented afterwards.

**2.2. An efficient algorithm using multilinear finite elements.** We define the hat function  $\Lambda_z^\mu$  with center  $z \in \mathbb{R}$  and size parameter  $\mu > 0$  by:

$$(2.10) \quad \Lambda_z^\mu(y) = \frac{y - (z - \mu)}{\mu} \mathbb{1}_{y \in [z - \mu, z]} + \frac{z + \mu - y}{\mu} \mathbb{1}_{y \in [z, z + \mu]}.$$

Observe that  $\mathbb{E}(\Lambda_z^\mu(G_1))$  is known in explicit form when  $G_1$  is a scalar Gaussian r.v. (like the proxy): therefore, replacing  $h$  by a linear interpolation  $\hat{h}$  (using the  $\Lambda_z^\mu$ -function) leads to a fully explicit formula for (2.6). To extend to the  $d$ -dimensional case, we wish to use tensor products of such a function basis in all directions to provide an interpolation of  $h$ . However remark that for any  $(G_1, G_2)$  Gaussian vector and any  $z_1, z_2, \mu_1, \mu_2$ , the calculus of  $\mathbb{E}[\Lambda_{z_1}^{\mu_1}(G_1)\Lambda_{z_2}^{\mu_2}(G_2)]$  is not tractable, except in the case of *zero correlation*; thus an additional ingredient is necessary to maintain explicit formulas. In order to be placed in a situation

of uncorrelated Gaussian r.v., we introduce an affine transformation  $\mathcal{A}$  of the space, composed of a rotation using the  $d$ -dimensional diagonal matrix  $\mathcal{U}_V$  (involved in the diagonal decomposition of  $\mathcal{V}_T^P$ ) and a translation of vector  $m_T^P$  (the expectation of  $X_T^P$ ). The following presentation is aimed at providing the construction of the right grid (nodes, directions, size).

▷ **Description of the methodology.** We consider a finite product grid in  $\mathbb{R}^d$  defined by:

(2.11)

$$\mathcal{Y} = (\mathcal{Y}_i^j)_{(i,j) \in \{1, \dots, d\} \times \{0, \dots, N\}}, \quad \mathcal{Y}_i^j = -R_i + j\delta_i, \quad R_i = R\lambda_i \sqrt{T}, \quad \delta_i = \delta\lambda_i \sqrt{T}, \quad \delta = \frac{2R}{N},$$

where we recall that  $\lambda_i^2 T$  are the eigenvalues of the covariance matrix  $\mathcal{V}_T^P$  and where the grid parameters  $R$  and  $\delta$  are to be specified according to the final approximation accuracy desired. We assume  $N \in \mathbb{N}^*$ . The grid  $\mathcal{Y}$  contains  $N^d$  small hypercubes and their vertices are the nodes with coordinates  $(y_1^{j_1}, \dots, y_d^{j_d})^*$  for any  $j_1, \dots, j_d \in \{0, \dots, N\}$ . Then we define a new grid  $X = (x^{j_1, \dots, j_d})_{j_1, \dots, j_d \in \{0, \dots, N\}}$  image of  $\mathcal{Y}$  by the transformation  $\mathcal{A} : x \mapsto \mathcal{A}x = m_T^P + \mathcal{U}_V x$ :

$$x^{j_1, \dots, j_d} = (x_1^{j_1, \dots, j_d}, \dots, x_d^{j_1, \dots, j_d})^* := \mathcal{A}(y_1^{j_1}, \dots, y_d^{j_d})^*.$$

The convex hull of  $\mathcal{Y}$  is the hypercube  $D^P$ , and let us introduce its image  $\tilde{D}^P$  by  $\mathcal{A}$ :

$$(2.12) \quad D^P = [-R_1, R_1] \times \dots \times [-R_d, R_d], \quad \tilde{D}^P := \mathcal{A}(D^P).$$

Then we define the multilinear interpolation of  $h$  based on the grid  $X$  by setting, for any  $x \in \mathbb{R}^d$ ,

$$(2.13) \quad h(x) \approx \hat{h}(x) := \sum_{j_1, \dots, j_d \in \{0, \dots, N\}} h(x^{j_1, \dots, j_d}) \prod_{i=1}^d \Lambda_{y_i^{j_i}}^{\delta_i} ((\mathcal{U}_V^{-1}(x - m_T^P))^i).$$

Notice that  $\hat{h}$  is continuous, vanishes outside the domain  $\mathcal{A}([-R_1 - \delta_1, R_1 + \delta_1] \times \dots \times [-R_d - \delta_d, R_d + \delta_d])$ , and the restriction of  $\hat{h}$  to  $D^P$  is Lipschitz continuous with a Lipschitz constant at most equal to  $C_{\text{Lip}, h}$ . The above construction is very similar to multilinear Lagrange finite elements on  $d$ -parallelotope, see [4] for a general reference.

▷ **Explicit approximation.** Using (2.13) and taking the expectation, we get for the leading order of the expansion (2.6):

$$(2.14) \quad \mathbb{E}[h(X_T^P)] \approx \mathbb{E}[\hat{h}(X_T^P)] = \sum_{j_1, \dots, j_d \in \{0, \dots, N\}} h(x^{j_1, \dots, j_d}) \mathbb{E} \left[ \prod_{i=1}^d \Lambda_{y_i^{j_i}}^{\delta_i} ((\mathcal{U}_V^{-1}(X_T^P - m_T^P))^i) \right],$$

where  $\mathcal{U}_V^{-1}(X_T^P - m_T^P)$  has the centered Gaussian law with independent components. Thus combining independence and scaling argument, setting

$$(2.15) \quad y_0 = (y_0^j)_{j \in \{-1, \dots, N+1\}} := (-R + j\delta)_{j \in \{-1, \dots, N+1\}}$$

and using (2.10)-(2.11)-(2.15), a straightforward calculus leads to

(2.16)

$$\mathbb{E} \left[ \prod_{i=1}^d \Lambda_{y_i^{j_i}}^{\delta_i} ((\mathcal{U}_V^{-1}(X_T^P - m_T^P))^i) \right] = \prod_{i=1}^d \mathbb{E} [\Lambda_{y_i^{j_i}}^{\delta_i} (\lambda_i \sqrt{T} W_1^i)] = \prod_{i=1}^d \mathbb{E} [\Lambda_{y_0^i}^{\delta_i} (W_1^i)] = \prod_{i=1}^d \beta_{j_i}^{\delta_i}(y_0),$$

where

(2.17)

$$\beta_j^{\delta}(y_0) := \frac{\beta(y_0^{j+1}) - 2\beta(y_0^j) + \beta(y_0^{j-1}))}{\delta}, \quad \beta(x) := x\mathcal{N}(x) + \mathcal{N}'(x), \quad \mathcal{N}(x) = \int_{-\infty}^x \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy = \beta'(x).$$



Next for the corrective terms (2.7), we similarly replace  $h$  by  $\hat{h}$ , which gives  $\text{Cor}_{2,h} \approx \text{Cor}_{2,\hat{h}}$ . For any multi-index  $\alpha \in \{1, \dots, d\}^{|\alpha|}$ , we get for the derivatives of  $\hat{h}^P$  using (2.14)-(2.16):

$$\begin{aligned} \partial_{\epsilon_{\alpha_1}, \dots, \epsilon_{\alpha_{|\alpha|}}} \hat{h}^P(\epsilon) &= \sum_{j_1, \dots, j_d \in \{0, \dots, N\}} h(x^{j_1, \dots, j_d}) \partial_{\epsilon_{\alpha_1}, \dots, \epsilon_{\alpha_{|\alpha|}}}^{|\alpha|} \prod_{i=1}^d \mathbb{E}[\Lambda_{y_i}^{\delta_i}(\lambda_i \sqrt{T} W_1^1 + (\mathcal{U}_V^{-1} \epsilon)^i)], \\ \mathbb{E}[\Lambda_{y_i}^{\delta_i}(\lambda_i \sqrt{T} W_1^1 + (\mathcal{U}_V^{-1} \epsilon)^i)] &= \mathbb{E}[\Lambda_{y_0 - \frac{(\mathcal{U}_V^{-1} \epsilon)^i}{\lambda_i \sqrt{T}}}^{\delta} (W_1^1)] = \beta_{j_i}^{\delta} \left( y_0 - \frac{(\mathcal{U}_V^{-1} \epsilon)^i}{\lambda_i \sqrt{T}} \right). \end{aligned}$$

Thus it is sufficient to compute the perturbed coefficients  $\beta^{\delta}$  as in (2.17) according to the new translated grid  $y_0 - \frac{(\mathcal{U}_V^{-1} \epsilon)^i}{\lambda_i \sqrt{T}}$  and then to differentiate w.r.t.  $\epsilon$  at  $\epsilon = 0$ , which leads to explicit calculations. The next result summarizes the previous analysis, in combination with Theorem 2.1.

**THEOREM 2.4. (SAFE method with multilinear finite elements).**

Assume  $(\mathcal{H}_{x_0})$  and suppose that  $h \in \text{Lip}(\mathbb{R}^d, \mathbb{R})$ . Define

$$\bar{h}^P(\epsilon) := \sum_{j_1, \dots, j_d \in \{0, \dots, N\}} h(x^{j_1, \dots, j_d}) \prod_{i=1}^d \beta_{j_i}^{\delta} \left( y_0 - \frac{(\mathcal{U}_V^{-1} \epsilon)^i}{\lambda_i \sqrt{T}} \right).$$

where the weight functions  $\beta_{j_i}^{\delta}$  and the grid  $y_0$  are respectively defined in (2.17) and (2.15). Then we have

$$(2.18) \quad \mathbb{E}[h(X_T)] = \bar{h}^P(0) + \text{Cor}_{2,\hat{h}} + \text{Error}_{2,h}^{\text{SA}} + \text{Error}_h^{\text{FEL}},$$

where  $\text{Cor}_{2,\hat{h}}$  is obtained replacing  $\bar{h}^P$  by  $\hat{h}^P$  in  $\text{Cor}_{2,h}$  (defined in (2.7)) and where the error using the multilinear finite elements approximation is defined by:

$$(2.19) \quad \text{Error}_h^{\text{FEL}} := \bar{h}^P(0) - \hat{h}^P(0) + \text{Cor}_{2,h} - \text{Cor}_{2,\hat{h}}.$$

▷ **Accuracy results.** The accuracy of multilinear interpolation depends on the smoothness of the function  $h$  to approximate. Our goal is not to be exhaustive in this respect but rather to give few settings relevant for the practical applications that we have in mind. For a detailed exposure on Finite Elements accuracy, see [4]. We distinguish three kinds of increasingly strong assumptions:

**(H1):**  $h \in \text{Lip}(\mathbb{R}^d, \mathbb{R})$ .

**(H2):**  $h \in \text{Lip}(\mathbb{R}^d, \mathbb{R})$ , piecewise  $C^2$ , in the sense that there are an integer  $N_h \in \mathbb{N}^*$ ,  $N_h$  domains (non empty open connected sets of  $\mathbb{R}^d$ )  $(D^i)_{i \in \{1, \dots, N_h\}}$ , such that:

1.  $\forall i \in \{1, \dots, N_h\}$ , either the domain  $D^i$  has a compact boundary  $\partial D^i$  of class  $C^2$ , or  $D^i$  is a half-space,

$$2. \mathbb{R}^d = \bigcup_{i=1}^{i=N_h} \overline{D^i},$$

3.  $\forall i \in \{1, \dots, N_h\}$ , the restriction of  $h$  to  $D^i$ , which is denoted by  $h_i$ , is a  $C^2(\overline{D^i}, \mathbb{R}^d)$ -function with bounded derivatives.

**(H3):**  $h \in C^2(\mathbb{R}^d, \mathbb{R})$  with bounded derivatives.

We state the accuracy results in the following Theorem, which proof is postponed to Section 4.

**THEOREM 2.5. (Accuracy of SAFE method with multilinear finite elements).**

Assume  $(\mathcal{H}_{x_0})$  and suppose that  $h$  satisfies at least **(H1)**. Recalling the density  $f^P$  of  $X_T^P$  in (1.4) and the domain  $\tilde{D}^P$  in (2.12), for any multi-index  $\alpha$  set

$$(2.20) \quad \mathcal{G}_h^{\alpha,T} = \int_{\mathbb{R}^d} \mathbb{1}_{y \notin \tilde{D}^P} h(y) \partial^\alpha (f^P(y - \epsilon)) \big|_{\epsilon=0} dy, \quad \mathcal{G}_h^{\alpha,I} = \int_{\mathbb{R}^d} \mathbb{1}_{y \in \tilde{D}^P} h(y) \partial^\alpha (f^P(y - \epsilon)) \big|_{\epsilon=0} dy.$$

Then, define  $\text{Cor}_{2,h}^T$  (respectively  $\text{Cor}_{2,h}^I$ ) replacing in  $\text{Cor}_{2,h}$  the sensitivities  $\partial^\alpha \bar{h}^P(0)$  by  $\mathcal{G}_h^{\alpha,T}$  (respectively  $\mathcal{G}_h^{\alpha,I}$ ) so that  $\text{Cor}_{2,h} = \text{Cor}_{2,h}^T + \text{Cor}_{2,h}^I$ ; proceed similarly with  $\mathcal{G}_{\hat{h}}^{\alpha,T}$ ,  $\mathcal{G}_{\hat{h}}^{\alpha,I}$ ,  $\text{Cor}_{2,\hat{h}}^T$  and  $\text{Cor}_{2,\hat{h}}^I$ . Then the multilinear finite elements error (2.19) is decomposed as

$$\text{Error}_h^{\text{FEL}} = \text{Error}_h^{\text{FEL},T} + \text{Error}_h^{\text{FEL},I},$$

where the Truncation Error

$$\text{Error}_h^{\text{FEL},T} := \mathbb{E}[(h(X_T^P) - \hat{h}(X_T^P)) \mathbb{1}_{X_T^P \notin \tilde{D}^P}] + \text{Cor}_{2,h}^T - \text{Cor}_{2,\hat{h}}^T$$

strongly depends on the size parameter  $R$  introduced in (2.11), and where the Interpolation Error on  $\tilde{D}^P$

$$\text{Error}_h^{\text{FEL},I} := \mathbb{E}[(h(X_T^P) - \hat{h}(X_T^P)) \mathbb{1}_{X_T^P \in \tilde{D}^P}] + \text{Cor}_{2,h}^I - \text{Cor}_{2,\hat{h}}^I$$

depends on the grid mesh  $\delta$ . On the one hand, for  $h \in \text{Lip}(\mathbb{R}^d, \mathbb{R})$ , the truncation error is such

$$(2.21) \quad |\text{Error}_h^{\text{FEL},T}| \leq c(|h(m_T^P)| + C_{\text{Lip},h}) \exp(-R^2/4).$$

On the other hand, the Interpolation Error is estimated as follows, according to the regularity of  $h$ :

$$(2.22) \quad \text{Error}_h^{\text{FEL},I} \leq c \begin{cases} C_{\text{Lip},h} \delta \mathcal{M}_0(\sigma, b) \sqrt{T} & \text{under (H1),} \\ \left\{ C_{\text{Lip},h} + \max_{i \leq N_h, \alpha: |\alpha|=2} |\partial^\alpha h_i|_\infty \right\} \delta \mathcal{M}_0(\sigma, b) \sqrt{T} [\delta + \mathcal{M}_0(\sigma, b) \sqrt{T}] & \text{under (H2),} \\ \sup_{\alpha: |\alpha|=2} |\partial^\alpha h|_\infty \delta^2 [\mathcal{M}_0(\sigma, b) \sqrt{T}]^2 & \text{under (H3),} \end{cases}$$

where the generic constant  $c$  in case **(H2)** depends on the domains.

**2.3. Final approximation and complexity of the SAFE algorithm based on multilinear finite elements.** Combining Theorems 2.1 (weak approximation with the Gaussian proxy) and 2.5 (suitable interpolation of  $h$  using the hat functions), we derive a final approximation of  $\mathbb{E}[h(X_T)]$  with a choice of parameters  $R$  and  $\delta$  (see (2.11)) allowing to obtain a global error of order at most equal to

$$\mathcal{E} = [\mathcal{M}_0(\sigma, b) \sqrt{T}]^3.$$

The proof of the following Theorem is left to the reader.

**THEOREM 2.6.** Assume  $(\mathcal{H}_{x_0})$  and suppose that  $h$  satisfies at least **(H1)**. Consider the local approximation  $\hat{h}$  of  $h$  defined in (2.13) with parameters  $R$  and  $\delta$  set as follows:

$$R := 2 \sqrt{\log(1/\mathcal{E})}, \quad \delta := c \begin{cases} [\max_i \lambda_i \sqrt{T}]^2 & \text{under (H1),} \\ \max_i \lambda_i \sqrt{T} & \text{under (H2),} \\ [\max_i \lambda_i \sqrt{T}]^{\frac{1}{2}} & \text{under (H3),} \end{cases}$$

for an arbitrary fixed constant  $c$ . Then, the global error is of order 3 w.r.t.  $\mathcal{M}_0(\sigma, b) \sqrt{T}$ :

$$\mathbb{E}[h(X_T)] = \mathbb{E}[\hat{h}(X_T^P)] + \text{Cor}_{2,\hat{h}} + \mathcal{O}([\mathcal{M}_0(\sigma, b) \sqrt{T}]^3).$$

Now, let us analyze the algorithm complexity w.r.t. the target error  $\mathcal{E}$ , according to the regularity of  $h$ . Denote by  $C(d)$  the computational cost for the elementary operations at each node  $x^{j_1, \dots, j_d}$  of the local approximation: apart from the evaluation of  $h(x^{j_1, \dots, j_d})$ , computations are mainly dedicated to the calculus of the  $\beta$  weights defined in (2.17) and their derivatives, which is simple and can even be made off-line. Therefore, the total computational cost of the algorithm is  $C_{\text{calculus}}^{\text{FEL}} = \mathcal{O}(C(d)(N+1)^d)$ . Since  $N = 2R/\delta$ , the complexity of the algorithm to reach the target error  $\mathcal{E} = [\mathcal{M}_0(\sigma, b) \sqrt{T}]^3$  can be evaluated in the following manner.

**COROLLARY 2.7.** *With the previous notations and assumptions, as  $\mathcal{E} \rightarrow 0$  we have*

$$(2.23) \quad C_{\text{calculus}}^{\text{FEL}} = \begin{cases} \mathcal{O}([\log(1/\mathcal{E})]^{d/2} \mathcal{E}^{-\frac{2d}{3}}) & \text{under (H1),} \\ \mathcal{O}([\log(1/\mathcal{E})]^{d/2} \mathcal{E}^{-\frac{d}{3}}) & \text{under (H2),} \\ \mathcal{O}([\log(1/\mathcal{E})]^{d/2} \mathcal{E}^{-\frac{d}{6}}) & \text{under (H3).} \end{cases}$$

Let us briefly discuss the theoretical efficiency of our algorithm in comparison to a direct Monte-Carlo method. If we perform  $M$  simulations of the diffusion  $X_T$  via an Euler scheme  $X_T^{\Delta t}$  with time step  $\Delta t$ , the total computational cost is of order  $M \times (T/\Delta t)$ , whereas the mean square error is (see [6])

$$\text{Var}[h(X_T^{\Delta t})]M^{-1} + \left( \mathbb{E}[h(X_T)] - \mathbb{E}[h(X_T^{\Delta t})] \right)^2.$$

The first term (statistical error) is approximately equal to

$$\text{Var}[h(X_T)]M^{-1} = \text{Var}[h(X_T) - h(x_0)]M^{-1} = \mathcal{O}([C_{\text{Lip},h} \mathcal{M}_0(\sigma, b) \sqrt{T}]^2 M^{-1}).$$

The second error term (discretization error) is a bit delicate to analyze under our assumptions: Kebaier shows in [12, Proposition 2.2] that for any  $\alpha \in (\frac{1}{2}, 1]$ , there is a Lipschitz function  $h$  so that the discretization error is  $\mathcal{O}((\Delta t)^\alpha)$ . In [21, 2], the order  $\alpha = 1$  is established for smooth  $h$  or for uniform (hypo)-elliptic  $\sigma$ , but these assumptions are not fulfilled in our setting. To encompass general results, we rather use strong convergence estimates, and we specialize them to our setting:

$$|\mathbb{E}[h(X_T)] - \mathbb{E}[h(X_T^{\Delta t})]| \leq C_{\text{Lip},h} \mathbb{E}|X_T - X_T^{\Delta t}| \leq_c C_{\text{Lip},h} [\mathcal{M}_0(\sigma, b)]^2 \sqrt{T} \sqrt{\Delta t}.$$

We now tune the parameters to achieve a  $L_2$ -error of the same order as the SAFE method when  $\mathcal{E} \rightarrow 0$ , i.e. to have a mean square error  $\mathcal{E}^2 = [\mathcal{M}_0(\sigma, b) \sqrt{T}]^6$ : this is achieved by taking  $M^{-1} \sim [\mathcal{M}_0(\sigma, b) \sqrt{T}]^4 = \mathcal{E}^{\frac{4}{3}}$  and  $\Delta t \sim [\mathcal{M}_0(\sigma, b)]^2 T^2 = \mathcal{E}^{\frac{2}{3}} T$ . Therefore, the computational cost is  $C_{\text{calculus}}^{\text{MC}} = \mathcal{O}(\mathcal{E}^{-2})$ , independently of the dimension. Thus, in view of (2.23) and neglecting logarithmic factors, the SAFE method with multilinear finite elements is (in the sense of this theoretical comparison) more competitive than Monte-Carlo methods up to the dimension

$$d = 3 \text{ under (H1),} \quad d = 6 \text{ under (H2),} \quad d = 12 \text{ under (H3).}$$

Subsequent experiments are coherent with these rules.

**2.4. SAFE with multiquadratic finite elements.** For smooth functions at least of class  $C^3(\mathbb{R}^d, \mathbb{R})$ , we propose an extension using multiquadratic finite elements [4, Section 3.5] allowing to reach a given accuracy with a lower computational cost. We define two basis functions of center and size parameters  $z \in \mathbb{R}$  and  $\mu > 0$ :

$$\begin{aligned}\zeta_z^\mu(y) &= \frac{y - (z - \mu)}{\mu} \left( 2 \left( \frac{y - (z - \mu)}{\mu} \right) - 1 \right) \mathbb{1}_{y \in [z - \mu, z]} + \frac{z + \mu - y}{\mu} \left( 2 \left( \frac{z + \mu - y}{\mu} \right) - 1 \right) \mathbb{1}_{y \in [z, z + \mu]}, \\ \Xi_z^\mu(y) &= \frac{(y - (z - \mu))}{\mu} \frac{(z + \mu - y)}{\mu} \mathbb{1}_{y \in [z - \mu, z + \mu]}.\end{aligned}$$

To obtain a multiquadratic interpolation of  $h$ , for any  $i \in \{1, \dots, d\}$  and any  $j \in \{0, \dots, N - 1\}$  we denote by  $y_0^{j+\frac{1}{2}} = -R + (j + \frac{1}{2})\delta$  and  $y_i^{j+\frac{1}{2}} = -R_i + (j + \frac{1}{2})\delta_i$  the middles of respectively the segments  $[y_0^j, y_0^{j+1}]$  and  $[y_i^j, y_i^{j+1}]$ . Then we naturally extend  $X = (x^{j_1, \dots, j_d})_{j_1, \dots, j_d \in \{0, \frac{1}{2}, \dots, N - \frac{1}{2}, N\}} = (\mathcal{A}(y_1^{j_1}, \dots, y_d^{j_d}))_{j_1, \dots, j_d \in \{0, \frac{1}{2}, \dots, N - \frac{1}{2}, N\}}$  and the multiquadratic interpolation of  $h$  is defined by:

$$\begin{aligned}(2.24) \quad Qh(x) &:= \sum_{j_1, \dots, j_d \in \{0, \frac{1}{2}, \dots, N - \frac{1}{2}, N\}} h(x^{j_1, \dots, j_d}) \prod_{i=1}^d \left[ \zeta_{y_i^{j_i}}^{\delta_i} ((\mathcal{U}_V^{-1}(x - m_T^P))^i) \mathbb{1}_{2j_i \equiv 0[2]} \right. \\ &\quad \left. + \Xi_{y_i^{j_i}}^{\delta_i/2} ((\mathcal{U}_V^{-1}(x - m_T^P))^i) \mathbb{1}_{2j_i \equiv 1[2]} \right],\end{aligned}$$

which is a continuous approximation, piecewise quadratic in all directions on  $\tilde{D}^P$  and vanishing outside the domain  $\mathcal{A}([-R_1 - \delta_1, R_1 + \delta_1] \times \dots \times [-R_d - \delta_d, R_d + \delta_d])$ . A direct computation (a little bit tedious but without mathematical difficulty) yields:

$$\begin{aligned}\mathbb{E}[\zeta_{y_i^{j_i}}^{\delta_i} ((\mathcal{U}_V^{-1}(X_T^P - m_T^P))^i)] &= 4\mathcal{B}_{j_i}^\delta(y_0) - \beta_{j_i}^\delta(y_0), \\ \mathbb{E}[\Xi_{y_i^{j_i}}^{\delta_i/2} ((\mathcal{U}_V^{-1}(X_T^P - m_T^P))^i)] &= -2\mathcal{B}_{j_i}^{\delta/2}(y_0) + 2\beta_{j_i}^{\delta/2}(y_0),\end{aligned}$$

where the function  $\beta$  is defined in (2.17) and where:

$$\begin{aligned}(2.25) \quad \mathcal{B}_j^\delta(y_0) &:= \frac{\mathcal{B}(y_0^{j+1}) - \mathcal{B}(y_0^j) - 2\delta\mathcal{B}'(y_0^j)}{\delta^2}, \quad \mathcal{B}(x) := \int_{-\infty}^x \beta(u) du = \frac{(x^2 + 1)\mathcal{N}(x) + x\mathcal{N}'(x)}{2}, \\ \beta_j^{\delta/2}(y_0) &:= \frac{\beta(y_0^{j+1/2}) - 2\beta(y_0^j) + \beta(y_0^{j-1/2})}{\delta/2}, \quad \mathcal{B}_j^{\delta/2}(y_0) := \frac{\mathcal{B}(y_0^{j+1/2}) - \mathcal{B}(y_0^{j-1/2}) - 2(\delta/2)\mathcal{B}'(y_0^j)}{(\delta/2)^2}.\end{aligned}$$

We are now in a position to announce the following Theorem, which proof is very similar to those of Theorems 2.4-2.5 and is thus left to the reader.

**THEOREM 2.8. (SAFE using multiquadratic finite elements).**

Assume  $(\mathcal{H}_{x_0})$  and suppose that  $h \in C^3(\mathbb{R}^d, \mathbb{R})$  with bounded derivatives. Define

$$\begin{aligned}\overline{Qh}^P(\epsilon) &:= \sum_{j_1, \dots, j_d \in \{0, \frac{1}{2}, \dots, N - \frac{1}{2}, N\}} h(x^{j_1, \dots, j_d}) \prod_{i=1}^d \left[ \left( 4\mathcal{B}_{j_i}^\delta(y_0 - \frac{(\mathcal{U}_V^{-1}\epsilon)^i}{\lambda_i \sqrt{T}}) - \beta_{j_i}^\delta(y_0 - \frac{(\mathcal{U}_V^{-1}\epsilon)^i}{\lambda_i \sqrt{T}}) \right) \mathbb{1}_{2j_i \equiv 0[2]} \right. \\ &\quad \left. + \left( -2\mathcal{B}_{j_i}^{\delta/2}(y_0 - \frac{(\mathcal{U}_V^{-1}\epsilon)^i}{\lambda_i \sqrt{T}}) + 2\beta_{j_i}^{\delta/2}(y_0 - \frac{(\mathcal{U}_V^{-1}\epsilon)^i}{\lambda_i \sqrt{T}}) \right) \mathbb{1}_{2j_i \equiv 1[2]} \right],\end{aligned}$$

where the weight functions  $\mathcal{B}_{j_i}^\delta$ ,  $\mathcal{B}_{j_i}^{\delta/2}$ ,  $\beta_{j_i}^\delta$  and  $\beta_{j_i}^{\delta/2}$  and the grid  $y_0$  are defined in (2.25), (2.17) and (2.15). Then

$$\mathbb{E}[h(X_T)] = \overline{Qh}^P(0) + \text{Cor}_{2, Qh} + O(C_{\text{Lip}, h} \mathcal{M}_1(\sigma, b) [\mathcal{M}_0(\sigma, b)]^2 T^{\frac{3}{2}}) + \text{Error}_h^{\text{FEQ}, T} + \text{Error}_h^{\text{FEQ}, I},$$

where the truncation error  $\text{Error}_h^{\text{FEQ},T}$  is estimated as in (2.21) and where the interpolation error using the multiquadratic interpolation is estimated as follows:

$$\text{Error}_h^{\text{FEQ},I} = O(\delta^3 [\mathcal{M}_0(\sigma, b) \sqrt{T}]^3).$$

In order to achieve a global accuracy  $\mathcal{E} = [\mathcal{M}_0(\sigma, b) \sqrt{T}]^3$ , we conclude as for Corollary 2.7: namely, take  $R = 2\sqrt{\log(1/\mathcal{E})}$  and  $\delta = c$  for an arbitrary fixed constant  $c > 0$ . It leads to the following complexity result, which demonstrates a better asymptotic theoretical efficiency compared to Monte-Carlo methods in any dimension.

**COROLLARY 2.9.** *With the previous notations and assumptions, as  $\mathcal{E} \rightarrow 0$  we have  $C_{\text{calculus}}^{\text{FEQ}} = O([\log(1/\mathcal{E})]^{d/2})$ .*

For functions mixing various local regularity properties, we have better to use different shape functions and different mesh sizes, and possibly sparse grids [5]. It can be done easily according to the examples to handle, we do not develop further this adaptive viewpoint.

**3. Proof of the error estimate in Theorem 2.1.** The estimate of  $\text{Error}_{2,h}^{\text{SA}}$  provided in (2.8) is proved in three steps:

1.  $L^p$ -estimates of the interpolated process  $X^\eta$  and of its derivatives;
2. Gaussian regularization of  $h$  with a small Brownian perturbation;
3. Malliavin integration by parts formula.

We may stress that the step 2 with Gaussian regularization is a crucial ingredient for a secured use of Malliavin calculus integration by parts under the sole pointwise ellipticity  $(\mathcal{H}_{x_0})$ .

**3.1.  $L^p$ -norm estimates of  $X^\eta - x_0$ ,  $\dot{X}^\eta$  and  $\ddot{X}^\eta$ .** **LEMMA 3.1.** *Assume  $(\mathcal{H}_{x_0})$ . We have the following estimates  $\forall p \geq 1$ :*

(3.1)

$$\sup_{t \in [0, T], \eta \in [0, 1]} \|X_t^\eta - x_0\|_p \leq_c \mathcal{M}_0(\sigma, b) \sqrt{T},$$

(3.2)

$$\sup_{t \in [0, T], \eta \in [0, 1]} \|\dot{X}_t^\eta\|_p \leq_c \mathcal{M}_1(\sigma, b) \mathcal{M}_0(\sigma, b) T, \quad \sup_{t \in [0, T], \eta \in [0, 1]} \|\ddot{X}_t^\eta\|_p \leq_c \mathcal{M}_1(\sigma, b) [\mathcal{M}_0(\sigma, b)]^2 T^{\frac{3}{2}}.$$

*Proof.* W.l.o.g. we can assume  $p \geq 2$ . The estimate (3.1) is standard using classic inequalities and the upper bounds on  $b$  and  $\sigma$  in  $(\mathcal{H}_{x_0})$ . Regarding the estimate of  $\dot{X}^\eta$ , start from (2.2): usual computations based on Burkholder-Davis-Gundy inequalities and the Gronwall lemma lead to

$$\mathbb{E}|\dot{X}_t^\eta|^p \leq_c (\mathcal{M}_1(\sigma, b) \sqrt{T})^p \sup_{t \in [0, T]} \mathbb{E}|X_t^\eta - x_0|^p,$$

with some generic constant  $c$  uniform in  $t \in [0, T]$  and  $\eta \in [0, 1]$ . Then plug the estimate (3.1) into this inequality to conclude. Finally for the last estimate of (3.2), start from (2.3) and apply the same inequalities to obtain:

$$\mathbb{E}|\ddot{X}_t^\eta|^p \leq_c (\mathcal{M}_1(\sigma, b) \sqrt{T})^p \left\{ \sup_{t \in [0, T]} \mathbb{E}|X_t^\eta - x_0|^{2p} + \sup_{t \in [0, T]} \mathbb{E}|\dot{X}_t^\eta|^{2p} + \sup_{t \in [0, T]} \mathbb{E}|\dot{X}_t^\eta|^p \right\};$$

using the previous estimates easily achieves the proof.  $\square$

**3.2. Regularization of  $h$  with a small noise perturbation.** Let  $\bar{W}$  be an extra independent  $d$ -dimensional Brownian motion defined on the same probability space, suppose that  $\mathcal{M}_1(\sigma, b) \neq 0$  w.l.o.g. and define the small parameter  $\xi = \mathcal{M}_1(\sigma, b) [\mathcal{M}_0(\sigma, b)]^2 T > 0$  and the  $C^\infty(\mathbb{R}^d, \mathbb{R})$ -function :

$$(3.3) \quad h_\xi(x) = \mathbb{E}[h(x + \xi \bar{W}_T)] = \mathbb{E}[h_{\xi/\sqrt{2}}(x + \xi \bar{W}_{\frac{T}{2}})].$$

Replacing  $h$  by  $h_\xi$  in our expansion analysis (2.6) induces extra errors quantified below.

LEMMA 3.2. *Assume  $(\mathcal{H}_{x_0})$ . For any  $h \in \text{Lip}(\mathbb{R}^d, \mathbb{R})$  and any multi-index  $\alpha$ , we have:*

$$\begin{aligned} |h - h_\xi|_\infty &\leq c C_{\text{Lip},h} \mathcal{M}_1(\sigma, b) [\mathcal{M}_0(\sigma, b)]^2 T^{\frac{3}{2}}, \\ |\partial^\alpha \bar{h}^P(0) - \partial^\alpha \bar{h}_\xi^P(0)| &\leq c C_{\text{Lip},h} \mathcal{M}_1(\sigma, b) [\mathcal{M}_0(\sigma, b)]^2 T^{\frac{3}{2}} (\mathcal{M}_0(\sigma, b) \sqrt{T})^{-|\alpha|}. \end{aligned}$$

*Proof.* The first estimate readily follows from (3.3) and  $h \in \text{Lip}(\mathbb{R}^d, \mathbb{R})$ . About the second estimate, using the density  $f^P$  of  $X_T^P$  (see (1.4)), write

$$\partial^\alpha \bar{h}^P(0) - \partial^\alpha \bar{h}_\xi^P(0) = \int_{\mathbb{R}^d} [h(y) - h_\xi(y)] \partial^\alpha (f^P(y - \epsilon)) \Big|_{\epsilon=0} dy,$$

and complete the proof by combining the first estimate with standard upper bounds for the derivatives of  $f^P$  (see Property 1.2).  $\square$

An important consequence of the above lemma is a nice control of correction and error terms w.r.t.  $h$  and  $h_\xi$ . The proof is straightforward and we skip it.

COROLLARY 3.3. *Under the assumptions of Lemma 3.2, we have*

$$\begin{aligned} |\text{Cor}_{2,h} - \text{Cor}_{2,h_\xi}| &\leq c [\mathcal{M}_1(\sigma, b)]^2 [\mathcal{M}_0(\sigma, b)]^2 T^2, \\ |\text{Error}_{2,h}^{\text{SA}} - \text{Error}_{2,h_\xi}^{\text{SA}}| &:= |\mathbb{E}[h(X_T)] - \mathbb{E}[h(X_T^P)] - \text{Cor}_{2,h} - (\mathbb{E}[h_\xi(X_T)] - \mathbb{E}[h_\xi(X_T^P)] - \text{Cor}_{2,h_\xi})| \\ &\leq c \mathcal{M}_1(\sigma, b) [\mathcal{M}_0(\sigma, b)]^2 T^{\frac{3}{2}}. \end{aligned}$$

Hence, proving the error estimate (2.8) is reduced to prove the following.

PROPOSITION 3.4. *Under previous assumptions, we have*

$$|\text{Error}_{2,h_\xi}^{\text{SA}}| \leq c C_{\text{Lip},h} \mathcal{M}_1(\sigma, b) [\mathcal{M}_0(\sigma, b)]^2 T^{\frac{3}{2}}.$$

**3.3. Malliavin integration by parts formula and proof of Proposition (3.4).** This step is aimed at deriving error estimates related to  $h_\xi$ , depending only on  $C_{\text{Lip},h_\xi} \leq C_{\text{Lip},h}$  and not on higher derivatives of  $h_\xi$ , thanks to Malliavin integration by parts formula: we make it possible because of the extra independent noise  $\xi \bar{W}_{T/2}$  in (3.3) which plays an important role on the degeneracy event (see Lemma 3.7). This ingredient is crucial to make our error analysis available.

Let us detail the argumentation. We consider the Malliavin calculus w.r.t the  $(q+d)$ -dimensional Brownian motion  $(W, \bar{W})$ , the Malliavin derivatives associated to  $W$  (respectively  $\bar{W}$ ) being denoted by  $D$  (respectively  $\bar{D}$ ). We refer to [20] for the related theory and the notations. First we extend the estimates provided in Lemma 3.1 to the norms  $\|\cdot\|_{k,p}$  related to the Sobolev space  $\mathbb{D}^{k,p}$ ,  $k = 1, 2$ .

LEMMA 3.5. *Under  $(\mathcal{H}_{x_0})$ , for any  $\eta \in [0, 1]$ ,  $i \in \{1, \dots, d\}$  and  $t \in [0, T]$ ,  $(X_t^\eta)^i \in \mathbb{D}^{2,\infty}$  and  $(\dot{X}_t^\eta)^i \in \mathbb{D}^{1,\infty}$ . In addition we have the following estimates for any  $p \geq 1$ :*

$$(3.4) \quad \|D_r(X_t^\eta)^i\|_p \leq c \mathcal{M}_0(\sigma, b), \quad \|D_{r,s}^2(X_t^\eta)^i\|_p \leq c \mathcal{M}_0(\sigma, b) \mathcal{M}_1(\sigma, b),$$

$$(3.5) \quad \|D_r(\dot{X}_t^\eta)^i\|_p \leq c \mathcal{M}_1(\sigma, b) \mathcal{M}_0(\sigma, b) \sqrt{T},$$

uniformly in  $r, s, t \in [0, T]$ ,  $i \in \{1, \dots, d\}$  and  $\eta \in [0, 1]$ .

*Proof.* The inclusions in  $\mathbb{D}^{1,\infty}$  and  $\mathbb{D}^{2,\infty}$  are standard to justify under our assumptions, we skip it and we focus on the  $L_p$ -estimates. W.l.o.g. we assume  $p \geq 2$ . We only detail the computations for  $\dot{X}_t^\eta$ . Start from (2.2) to get  $\forall t \in [0, T]$ ,  $\forall r \in [0, t]$ ,  $\forall (i, k) \in \{1, \dots, d\} \times$

$\{1, \dots, q\}$ :

$$\begin{aligned}
(D_r \dot{X}_t^\eta)_k^i &= \nabla \sigma_k^i(r, x_0 + \eta(X_r^\eta - x_0))(X_r^\eta - x_0 + \eta \dot{X}_r^\eta) \\
&+ \sum_{j=1}^q \int_r^t (X_u^\eta - x_0 + \eta \dot{X}_u^\eta)^* H(\sigma_j^i)(u, x_0 + \eta(X_u^\eta - x_0))(D_r X_u^\eta)_k dW_u^j \\
&+ \sum_{j=1}^q \int_r^t \nabla \sigma_j^i(u, x_0 + \eta(X_u^\eta - x_0))(D_r X_u^\eta + \eta D_r \dot{X}_u^\eta)_k dW_u^j \\
&+ \int_r^t (X_u^\eta - x_0 + \eta \dot{X}_u^\eta)^* H(b^i)(u, x_0 + \eta(X_u^\eta - x_0))(D_r X_u^\eta)_k du \\
&+ \int_r^t \nabla b^i(u, x_0 + \eta(X_u^\eta - x_0))(D_r X_u^\eta + \eta D_r \dot{X}_u^\eta)_k du.
\end{aligned}$$

Then apply the Young, Burkholder-Davis-Gundy and Hölder inequalities combined to the Gronwall lemma: it gives

$$\begin{aligned}
\mathbb{E}|(D_r \dot{X}_t^\eta)_k|^p &\leq_c [\mathcal{M}_1(\sigma, b)]^p \left\{ \mathbb{E}|X_r^\eta - x_0|^p + \mathbb{E}|\dot{X}_r^\eta|^p + \sup_{u \in [0, T]} \sqrt{\mathbb{E}|(D_r X_u^\eta)_k|^{2p}} T^{p/2} \right. \\
&\quad \times \left( \sup_{u \in [0, T]} \sqrt{\mathbb{E}|X_u^\eta - x_0|^{2p}} + \sqrt{\sup_{u \in [0, T]} \mathbb{E}|\dot{X}_u^\eta|^{2p}} \right) + \sup_{u \in [0, T]} \mathbb{E}|(D_r X_u^\eta)_k|^p T^{p/2} \Big\}.
\end{aligned}$$

In view of (3.1), (3.2) and (3.4), the announced result (3.5) is proved.  $\square$

We now state the key result to establish Proposition (3.4).

**PROPOSITION 3.6.** *Assume  $(\mathcal{H}_{x_0})$ . For  $\eta \in [0, 1]$ , we consider the  $d$ -dimensional random variable:*

$$(3.6) \quad F^\eta = X_T^P + \eta(X_T - X_T^P) + \xi \bar{W}_{\frac{T}{2}}.$$

Then for any  $Y \in \mathbb{D}^{1, \infty}$  and any  $i, j \in \{1, \dots, d\}$ , there exists a random variable  $Y_{j, \eta} \in \cap_{p \geq 1} L^p$  such that

$$\mathbb{E}[\partial_{x_i, x_j}^2 h_{\xi/\sqrt{2}}(F_\eta) Y] = \mathbb{E}[\partial_{x_i} h_{\xi/\sqrt{2}}(F_\eta) Y_{j, \eta}],$$

where  $\|Y_{j, \eta}\|_p \leq_c \|Y\|_{1, 3p} (\mathcal{M}_0(\sigma, b) \sqrt{T})^{-1}$ , uniformly in  $\eta \in [0, 1]$ , for any  $p \geq 1$ .

Before proving it, let us complete the proof of Proposition 3.4. Define the residual processes

$$(3.7) \quad R_T^{0, X} := X_T - X_T^P = \int_0^1 \dot{X}_T^\eta d\eta, \quad R_T^{1, X} := X_T - X_T^P - \dot{X}_T = \int_0^1 (1 - \eta) \ddot{X}_T^\eta d\eta,$$

which enable us to represent the error (2.5) as follows:

$$(3.8) \quad \text{Error}_{2, h_\xi}^{\text{SA}} = \mathbb{E}[\nabla h_\xi(X_T^P) R_T^{1, X}] + \mathbb{E}\left(\int_0^1 (1 - \eta) [R_T^{0, X}]^* H(h_\xi)(X_T^P + \eta(X_T - X_T^P)) R_T^{0, X} d\eta\right).$$

The first error term is easily handled combining (3.7), Lemma 3.1 and  $C_{\text{Lip}, h_\xi} \leq C_{\text{Lip}, h}$ :

$$\mathbb{E}[\nabla h_\xi(X_T^P) R_T^{1, X}] = O(C_{\text{Lip}, h}(\mathcal{M}_1(\sigma, b) [\mathcal{M}_0(\sigma, b)]^2 T^{\frac{3}{2}}).$$

For the second error term of (3.8), use (3.3), apply for any  $i, j \in \{1, \dots, d\}$  the above Proposition 3.6 with  $Y = (R_{0,t}^X)_i (R_{0,t}^X)_j$  and use Lemmas 3.1 and 3.5 combined with (3.7) to get:

$$\begin{aligned} & \mathbb{E} \left[ \int_0^1 (1-\eta) [R_T^{0,X}]^* H(h_\xi) (X_T^P + \eta(X_T - X_T^P)) R_T^{0,X} d\eta \right] \\ &= \int_0^1 (1-\eta) \mathbb{E} \left[ [R_T^{0,X}]^* H(h_{\xi/\sqrt{2}}) (F^\eta) R_T^{0,X} \right] d\eta = O(C_{\text{Lip},h} [\mathcal{M}_1(\sigma, b)]^2 \mathcal{M}_0(\sigma, b) T^{\frac{3}{2}}). \end{aligned}$$

Proposition 3.4 is proved.

**3.4. Proof of Proposition 3.6.** It is clear that under  $(\mathcal{H}_{x_0})$ ,  $F^\eta$  defined in (3.6) is in  $\mathbb{D}^{2,\infty}$  and is non degenerate since its Malliavin covariance matrix  $\gamma_{F^\eta}$  is such that

$$(3.9) \quad \gamma_{F^\eta} = \int_0^T D_t(X_T^P + \eta(X_T - X_T^P)) [D_t(X_T^P + \eta(X_T - X_T^P))]^* dt + \xi^2 \frac{T}{2} I_d \geq \xi^2 \frac{T}{2} I_d,$$

$$(3.10) \quad \gamma_{F^0} = \mathcal{V}_T^P + \xi^2 \frac{T}{2} I_d \geq (\bar{C}_\eta)^{-2} [\mathcal{M}_0(\sigma, b)]^2 T I_d.$$

Then [20, Propositions 2.1.4 and 1.5.6] proves the existence of  $Y_{j,\eta}$  such that for any  $p \geq 1$ ,

$$\|Y_{j,\eta}\|_p \leq_c \|Y\|_{1,3p} \|\gamma_{F^\eta}^{-1}\|_{1,3p} \|(DF^\eta, \bar{D}F^\eta)\|_{1,3p} \leq_c \|Y\|_{1,3p} \|\gamma_{F^\eta}^{-1}\|_{1,3p} \|\mathcal{M}_0(\sigma, b) \sqrt{T}\|,$$

where we have used (3.4) and the value of  $\xi$ . To complete the proof of Proposition 3.6, apply the following estimates related to  $\gamma_{F^\eta}^{-1}$ :

LEMMA 3.7. Assume  $(\mathcal{H}_{x_0})$ . For any  $p \geq 1$  and any  $\eta \in [0, 1]$ , we have:

$$(3.11) \quad \|\det^{-1}(\gamma_{F^\eta})\|_p \leq_c (\mathcal{M}_0(\sigma, b) \sqrt{T})^{-2d},$$

$$(3.12) \quad \|(\gamma_{F^\eta}^{-1})_{ij}^i\|_p \leq_c (\mathcal{M}_0(\sigma, b) \sqrt{T})^{-2}, \quad \sup_{t \in [0, T]} \|D_t(\gamma_{F^\eta}^{-1})_{ij}^i\|_p \leq_c \mathcal{M}_1(\sigma, b) (\mathcal{M}_0(\sigma, b) \sqrt{T})^{-2},$$

uniformly in  $\eta \in [0, 1]$ .

*Proof.* Here we rely much on the assumption  $(\mathcal{H}_{x_0})$  about the oscillation of eigenvalues of  $\mathcal{V}_T^P$ . All the next generic constants are uniform in  $\eta \in [0, 1]$ . Using the definition of  $\gamma_{F^\eta}$  (3.9), write

$$\gamma_{F^\eta} - \gamma_{F^0} = \eta \int_0^T \left\{ D_t X_T^P [D_t(X_T - X_T^P)]^* + D_t(X_T - X_T^P) [D_t X_T^P + \eta D_t(X_T - X_T^P)]^* \right\} dt.$$

Then, combining (3.7), (3.4) and (3.5), it readily follows ( $\forall p \geq 1$ )

$$(3.13) \quad \|\gamma_{F^\eta} - \gamma_{F^0}\|_p \leq_c \mathcal{M}_1(\sigma, b) [\mathcal{M}_0(\sigma, b)]^2 T^{\frac{3}{2}}, \quad \|\gamma_{F^\eta}\|_p \leq_c (\mathcal{M}_0(\sigma, b) \sqrt{T})^2.$$

Thus, it is an easy exercise (using the Leibniz formula for determinant) to deduce

$$(3.14) \quad \frac{\|\det(\gamma_{F^\eta}) - \det(\gamma_{F^0})\|_p}{\det(\gamma_{F^0})} \leq_c \frac{\mathcal{M}_1(\sigma, b) [\mathcal{M}_0(\sigma, b)]^2 T^{\frac{3}{2}} (\mathcal{M}_0(\sigma, b) \sqrt{T})^{2(d-1)}}{(\mathcal{M}_0(\sigma, b) \sqrt{T})^{2d}} \leq \mathcal{M}_1(\sigma, b) \sqrt{T}.$$

We are in a position to prove (3.11). For any given  $p \geq 1$  and  $m \geq 1$ ,

$$\begin{aligned} & \mathbb{E}[\det^{-p}(\gamma_{F^\eta})] = \mathbb{E}[\det^{-p}(\gamma_{F^\eta}) \mathbb{1}_{\det(\gamma_{F^\eta}) \leq \frac{1}{2} \det(\gamma_{F^0})}] + \mathbb{E}[\det^{-p}(\gamma_{F^\eta}) \mathbb{1}_{\det(\gamma_{F^\eta}) > \frac{1}{2} \det(\gamma_{F^0})}] \\ & \leq_c (\xi^2 T)^{-dp} \mathbb{P}\left(\det(\gamma_{F^0}) - \det(\gamma_{F^\eta}) \geq \frac{\det(\gamma_{F^0})}{2}\right) + \det^{-p}(\gamma_{F^0}) \\ & \leq_c ([\mathcal{M}_1(\sigma, b)]^2 [\mathcal{M}_0(\sigma, b)]^4 T^3)^{-dp} \det^{-m}(\gamma_{F^0}) \mathbb{E}[|\det(\gamma_{F^\eta}) - \det(\gamma_{F^0})|^m] + (\mathcal{M}_0(\sigma, b) \sqrt{T})^{-2pd} \\ & \leq_c ([\mathcal{M}_1(\sigma, b)]^2 [\mathcal{M}_0(\sigma, b)]^4 T^3)^{-dp} (\mathcal{M}_1(\sigma, b) \sqrt{T})^m + (\mathcal{M}_0(\sigma, b) \sqrt{T})^{-2pd}, \end{aligned}$$



where we have notably used (3.9) at the first inequality, the Markov inequality, the definition of  $\xi$  and (3.10) at the second one and (3.14) at the last one. Then choose  $m = 4pd$  to readily obtain:

$$\mathbb{E}[\det^{-p}(\gamma_{F^n})] \leq_c (\mathcal{M}_0(\sigma, b) \sqrt{T})^{-2pd}.$$

The first estimate of (3.12) readily follows from (3.11)-(3.13). Regarding the second one, it is a consequence of  $D_t(\gamma_{F^n}^{-1})_j^i = -\sum_{m,l=1}^d (\gamma_{F^n}^{-1})_m^i (\gamma_{F^n}^{-1})_j^l D_l(\gamma_{F^n})_t^m$  (see [20, Lemma 2.1.6]) and of the estimate  $\|D_t(\gamma_{F^n})_j^i\|_p \leq_c \mathcal{M}_1(\sigma, b)[\mathcal{M}_0(\sigma, b)]^2 T$  which comes from (3.4).  $\square$

### 3.4. Proof of Proposition 3.6.

#### 4.1. Truncation Error $\text{Error}_h^{\text{FEL},T}$ .

$\triangleright$  **Estimate of  $\mathbb{E}[(h - \hat{h})(X_T^P) \mathbb{1}_{X_T^P \notin \tilde{D}^P}]$ .** By construction of  $\hat{h}$ , we have  $|\hat{h}(x)| \leq \sup_{|x| \in \tilde{D}^P} |h(x)| \leq |h(m_T^P)| + C_{\text{Lip},h} \sqrt{d} \max_i R_i$ , which becomes  $|\hat{h}(x)| \leq |h(m_T^P)| + \sqrt{d} \frac{\max_i R_i}{\min_i R_i} C_{\text{Lip},h} |x - m_T^P|$  for  $x \notin \tilde{D}^P$ . Besides,  $|h(x)| \leq |h(m_T^P)| + C_{\text{Lip},h} |x - m_T^P|$ . Therefore

$$(4.1) \quad |h(x) - \hat{h}(x)| \leq 2|h(m_T^P)| + \left(1 + \sqrt{d} \frac{\max_i R_i}{\min_i R_i}\right) C_{\text{Lip},h} |x - m_T^P|, \quad \forall x \notin \tilde{D}^P.$$

Then using the Cauchy-Schwarz inequality and Lemma 3.1, we get:

$$|\mathbb{E}[(h - \hat{h})(X_T^P) \mathbb{1}_{X_T^P \notin \tilde{D}^P}]| \leq_c |h(m_T^P)| \mathbb{P}(X_T^P \notin \tilde{D}^P) + C_{\text{Lip},h} \mathcal{M}_0(\sigma, b) \sqrt{T} \left[ \mathbb{P}(X_T^P \notin \tilde{D}^P) \right]^{\frac{1}{2}}.$$

By the definition (2.12) of the domains  $D^P$  and  $\tilde{D}^P$ , we have

$$(4.2) \quad \mathbb{P}(X_T^P \notin \tilde{D}^P) = \mathbb{P}(W_1 \notin [-R, R]^d) \leq d \mathbb{P}(|W_1^1| > R) \leq 2de^{-\frac{R^2}{2}}$$

using a standard Gaussian concentration inequality. Finally, we have proved

$$(4.3) \quad |\mathbb{E}[(h - \hat{h})(X_T^P) \mathbb{1}_{X_T^P \notin \tilde{D}^P}]| \leq_c (|h(m_T^P)| + C_{\text{Lip},h}) e^{-\frac{R^2}{4}}.$$

$\triangleright$  **Estimate of  $\text{Cor}_{2,h}^T - \text{Cor}_{2,\hat{h}}^T$ .** Starting from the definition (2.20) and using the inequality (4.1), the estimates (3.1)-(4.2) and the Hölder inequality, we derive that  $|\mathcal{G}_h^{\alpha,T} - \mathcal{G}_{\hat{h}}^{\alpha,T}|$  is bounded by

$$\begin{aligned} & \left\{ \int_{\mathbb{R}^d} \mathbb{1}_{y \notin \tilde{D}^P} |h(y) - \hat{h}(y)|^4 f^P(y) dy \right\}^{\frac{1}{4}} \left\{ \int_{\mathbb{R}^d} |(f^P(y))^{-1} \partial^\alpha (f^P(y - \epsilon))|_{\epsilon=0}|^4 f^P(y) dy \right\}^{\frac{1}{4}} \\ & \times \left\{ \int_{\mathbb{R}^d} \mathbb{1}_{y \notin \tilde{D}^P} f^P(y) dy \right\}^{\frac{1}{2}} \leq_c (|h(m_T^P)| + C_{\text{Lip},h}) (\mathcal{M}_0(\sigma, b) \sqrt{T})^{-|\alpha|} e^{-\frac{R^2}{4}}, \end{aligned}$$

which easily leads to  $|\text{Cor}_{2,h}^T - \text{Cor}_{2,\hat{h}}^T| \leq_c (|h(m_T^P)| + C_{\text{Lip},h}) e^{-\frac{R^2}{4}}$ . Combining this last estimate with (4.3) achieves the proof of (2.21).

**4.2. Interpolation Error  $\text{Error}_h^{\text{FEL},I}$ .** For any  $j_1, \dots, j_d \in \{0, \dots, N-1\}$  we denote by  $D^{P,j_1,\dots,j_d}$  the hyperrectangle  $[y_1^{j_1}, y_1^{j_1+1}] \times \dots \times [y_d^{j_d}, y_d^{j_d+1}]$  and we set  $\tilde{D}^{P,j_1,\dots,j_d} = \mathcal{A}(D^{P,j_1,\dots,j_d})$ .

$\triangleright$  **Estimate of  $\mathbb{E}[(h - \hat{h})(X_T^P) \mathbb{1}_{X_T^P \in \tilde{D}^P}]$ , cases (H1) and (H3).** Our local approximation consists in using a tensor product finite elements of order 1 on  $d$ -parallelotope. We have for any  $j_1, \dots, j_d \in \{0, \dots, N-1\}$

$$(4.4) \quad \begin{aligned} \text{diam}(\tilde{D}^{P,j_1,\dots,j_d}) &:= \max_{x, x' \in \tilde{D}^{P,j_1,\dots,j_d}} |x - x'| = \text{diam}(D^{P,j_1,\dots,j_d}) \\ &= \left( \sum_{i=1}^d \delta_i^2 \right)^{\frac{1}{2}} = \delta \left( \sum_{i=1}^d \lambda_i^2 T \right)^{\frac{1}{2}} \leq \mu := \bar{C}_V \sqrt{d} \delta \mathcal{M}_0(\sigma, b) \sqrt{T}. \end{aligned}$$

Hence an application of [4, Theorem 4.6.14] yields that

$$(4.5) \quad \sup_{j_1, \dots, j_d} \sup_{x \in \tilde{D}^{P, j_1 \dots j_d}} |h(x) - \hat{h}(x)| \leq c_0 \begin{cases} C_{\text{Lip}, h} \mu, & \text{case (H1)}, \\ \left( \sum_{\alpha: |\alpha|=2} |\partial^\alpha h|_\infty \right) \mu^2, & \text{case (H3)}, \end{cases}$$

where  $c_0$  is a universal constant. This obviously leads to :

$$(4.6) \quad |\mathbb{E}[(h - \hat{h})(X_T^P) \mathbb{1}_{X_T^P \in \tilde{D}^P}]| \leq c \begin{cases} C_{\text{Lip}, h} \delta \mathcal{M}_0(\sigma, b) \sqrt{T}, & \text{case (H1)}, \\ \left( \sum_{\alpha: |\alpha|=2} |\partial^\alpha h|_\infty \right) \delta^2 [\mathcal{M}_0(\sigma, b) \sqrt{T}]^2, & \text{case (H3)}. \end{cases}$$

▷ **Estimate of  $\mathbb{E}[(h - \hat{h})(X_T^P) \mathbb{1}_{X_T^P \in \tilde{D}^P}]$ , case (H2).** We denote by  $\mathcal{J}$  the set of integers  $\mathbf{j} = (j_1, \dots, j_d) \in \{0, \dots, N\}^d$  such that  $\tilde{D}^{P, j_1 \dots j_d}$  does not intersect  $\bigcup_{i=1}^{N_h} \partial D^i$ : for  $\mathbf{j} \in \mathcal{J}$ , the restriction of  $h$  to  $\tilde{D}^{P, j_1 \dots j_d}$  coincides with a  $C^2$ -function, to which we can apply the estimate (4.5) in the case (H3). Otherwise, we can nevertheless use the estimate in the case (H1). It gives

$$\begin{aligned} & \left| \mathbb{E}[(h - \hat{h})(X_T^P) \mathbb{1}_{X_T^P \in \tilde{D}^P}] \right| \\ & \leq c_0 \sum_{\mathbf{j} \in \mathcal{J}} \mathbb{E} \left( \mu^2 \max_{i \in \{1, \dots, N_h\}} \left( \sum_{\alpha: |\alpha|=2} |\partial^\alpha h_i|_\infty \right) \mathbb{1}_{X_T^P \in \tilde{D}^{P, j_1 \dots j_d}} \right) + c_0 \sum_{\mathbf{j} \notin \mathcal{J}} \mathbb{E} \left( \mu C_{\text{Lip}, h} \mathbb{1}_{X_T^P \in \tilde{D}^{P, j_1 \dots j_d}} \right). \end{aligned}$$

Moreover, by the definition of  $\mu$  (see (4.4)),  $x \in \tilde{D}^{P, j_1 \dots j_d}$  for some  $\mathbf{j} \notin \mathcal{J}$  implies that  $d(x, \partial D^i) \leq \mu$  for some  $i$ : therefore,  $\sum_{\mathbf{j} \notin \mathcal{J}} \mathbb{P}(X_T^P \in \tilde{D}^{P, j_1 \dots j_d}) \leq \sum_{i=1}^{N_h} \mathbb{P}(d(X_T^P, \partial D^i) \leq \mu)$ . An application of the following lemma finally leads to

$$(4.7) \quad \left| \mathbb{E}[(h - \hat{h})(X_T^P) \mathbb{1}_{X_T^P \in \tilde{D}^P}] \right| \leq c \left( C_{\text{Lip}, h} + \max_{i \in \{1, \dots, N_h\}} \left( \sum_{\alpha: |\alpha|=2} |\partial^\alpha h_i|_\infty \right) \right) \delta^2 \mathcal{M}_0(\sigma, b) \sqrt{T}.$$

LEMMA 4.1. Assume  $(\mathcal{H}_{x_0})$  and (H2), let  $\mu = \bar{C}_V \sqrt{\delta} \delta \mathcal{M}_0(\sigma, b) \sqrt{T}$ . Then  $\forall i \in \{1, \dots, N_h\}$ , we have:

$$\mathbb{P}(d(X_T^P, \partial D^i) \leq \mu) \leq c \delta.$$

*Proof.* We give here a proof in the particular case where  $D^i$  is a half-space of the form  $D^i = \{x \in \mathbb{R}^d, x_1 > 0\}$ . Using the bound (1.4) for the density of  $X_T^P$ , we have

$$\mathbb{P}(d(X_T^P, \partial D^i) \leq \mu) \leq c \int_{\mathbb{R}^d} \mathbb{1}_{-\mu < x_1 \leq \mu} \frac{e^{-\frac{|x - m_T^P|^2}{C_{0,d} [\mathcal{M}_0(\sigma, b)]^2 T}}}{(\mathcal{M}_0(\sigma, b) \sqrt{T})^d} dx \leq c \int_{-\mu}^{\mu} \frac{e^{-\frac{|x_1 - (m_T^P)_1|^2}{C_{0,d} [\mathcal{M}_0(\sigma, b)]^2 T}}}{(\mathcal{M}_0(\sigma, b) \sqrt{T})} dx_1,$$

the last integral being obviously bounded by  $\frac{\mu}{\mathcal{M}_0(\sigma, b) \sqrt{T}} \leq c \delta$ . For the general case, the idea is to locally map  $D^i$  into a half-space by using local charts, so that we are reduced to the first case. We skip these standard computations and refer for instance to [8]. □

▷ **Estimate of  $\text{Cor}_{2,h}^I - \text{Cor}_{2,\hat{h}}^I$ .** Using the definition (2.20), the estimates (4.5) of  $|\mathbb{1}_{y \in \tilde{D}^P}(h - \hat{h})(y)|$ , the upper bounds (1.4) on  $\partial^\alpha f^P$ , we easily get:

$$|\mathcal{G}_h^{\alpha, I} - \mathcal{G}_{\hat{h}}^{\alpha, I}| \leq c \begin{cases} C_{\text{Lip}, h} \delta (\mathcal{M}_0(\sigma, b) \sqrt{T})^{1-|\alpha|}, & \text{cases (H1) and (H2)}, \\ \left\{ \sum_{\alpha: |\alpha|=2} |\partial^\alpha h|_\infty \right\} \delta^2 (\mathcal{M}_0(\sigma, b) \sqrt{T})^{2-|\alpha|}, & \text{case (H3)}. \end{cases}$$

It readily follows that

$$|\text{Cor}_{2,h}^I - \text{Cor}_{2,\hat{h}}^I| \leq c \begin{cases} C_{\text{Lip}, h} \delta (\mathcal{M}_0(\sigma, b) \sqrt{T})^2, & \text{cases (H1) and (H2)}, \\ \left\{ \sum_{\alpha: |\alpha|=2} |\partial^\alpha h|_\infty \right\} \delta^2 (\mathcal{M}_0(\sigma, b) \sqrt{T})^3, & \text{case (H3)}. \end{cases}$$

Gathering the above estimate with (4.6)-(4.7) yields the announced result (2.22).

## 5. Numerical experiments.

**5.1. Model and set of parameters.** For the numerical tests, we consider the case of  $d$  independent one-dimensional elliptic diffusions  $(X^i)_{i \in \{1, \dots, d\}}$  driven by their own scalar Brownian motion ( $d = q$ ). We choose the same dynamics for all diffusions: for any  $i \in \{1, \dots, d\}$ , set

$$\sigma(x) = 1 - \frac{x}{1 + x + x^2}, \quad dX_t^i = \left(\mu + \frac{1}{2}\nu^2\sigma^{(1)}(X_t^i)\right)\sigma(X_t^i)dt + \nu\sigma(X_t^i)dW_t^i, \quad X_0^i = x_0.$$

We easily check that  $\sigma$  takes values in  $[\frac{2}{3}, 2]$  and is of class  $C^\infty$  with bounded derivatives, the first derivative vanishing at  $\pm 1$ . A key feature of this diffusion model is that (owing to the Lamperti transform, see [11, p. 294-295])  $X_t^i = g(f(x_0) + \mu t + \nu W_t^i)$  where  $g$  is the inverse function of

$$f(x) = \int_0^x \frac{dy}{\sigma(y)} = x + \frac{1}{2} \log(1 + x^2).$$

Thus, using numerical inversions of  $f$  to evaluate  $g$ , we can exactly simulate  $X_T$  without discretization scheme to get our reference values. We consider four types of terminal functions from the most regular case to the less regular one:

$$\begin{aligned} h_1(x) &= \frac{100e^{\frac{1}{d}\sum_{i=1}^d x_i}}{1 + e^{\frac{1}{d}\sum_{i=1}^d x_i}} \quad (C^\infty \text{ with bounded derivatives, case (H3)}), \\ h_2(x) &= 100e^{\frac{1}{d}\sum_{i=1}^d x_i} \quad (C^\infty, \text{ case (H3) with unbounded derivatives}), \\ h_3(x) &= \frac{100}{d} \left( \sum_{i=1}^d x_i \right)_+ \quad (\text{Lipschitz, case (H2)}), \\ h_4(x) &= 100 \max(x_1, \dots, x_d) \quad (\text{Lipschitz, case (H1)}). \end{aligned}$$

Besides, we investigate three sets of parameters:

- ( $\mathcal{P}_1$ ):  $x_0 = 0$ ,  $\nu = 20\%$ ,  $\mu = 0$  (standard situation);
- ( $\mathcal{P}_2$ ):  $x_0 = 0$ ,  $\nu = \mu = 20\%$  (large drift);
- ( $\mathcal{P}_3$ ):  $x_0 = 1$ ,  $\nu = 20\%$ ,  $\mu = 0$ , (small magnitude and variations of  $\sigma$ ).

In all the numerical tests, we set  $T = 1$  and the purpose is to compare the following approximations:

1. **SAFE Lin (H1)-(H2)-(H3)**: the SAFE method using multilinear finite elements (see Theorem 2.4) and fixing its parameters as follows (according to Theorems 2.5-2.6):  $R = 4$  and  $\delta = [\nu\sigma(x_0)]^2$ ,  $\delta = \nu\sigma(x_0)$  and  $\delta = [\nu\sigma(x_0)]^{\frac{1}{2}}$  for respectively **(H1)**-**(H2)**-**(H3)**.
2. **SAFE Quad 1-2-3-4-5**: the SAFE method using multiquadratic finite elements (see Theorem 2.8) using  $R = 4$  and  $\delta = 0.5$ ,  $\delta = 1$ ,  $\delta = 2$ ,  $\delta = 4$  and  $\delta = 8$  for respectively 1-2-3-4-5. The methods **SAFE Quad 4-5** using large steps are only investigated in dimension 10.
3. **MC**: the estimation by Monte-Carlo simulations without discretization scheme. Keep in mind that the execution times reported in the test tables should be multiplied by about a factor 100 to take into account the usual time discretization effort. In all the tests we use  $10^7$  sample paths. The value in parentheses is the half-width of the related 95%-symmetric confidence interval.
4. **MC Proxy**: the estimation by Monte-Carlo simulations of the stochastic approximation formula (2.6) (see Theorem (2.3)). In all the tests we use  $10^7$  sample paths.

TABLE 5.1

Estimation of the expectations in dimension 2 with MC, MC proxy, SAFE methods (set of parameters ( $\mathcal{P}_1$ )) and execution time.

method / function	$h_1$	$h_2$	$h_3$	$h_4$	exec. time
MC	49.47 ( $\pm 2.3\text{E}-2$ )	98.99 ( $\pm 9.0\text{E}-3$ )	4.89 ( $\pm 4.5\text{E}-3$ )	9.70 ( $\pm 9.5\text{E}-3$ )	1m2s
MC Proxy	49.49 ( $\pm 1.1\text{E}-2$ )	99.01 ( $\pm 2.3\text{E}-2$ )	4.84 ( $\pm 5.5\text{E}-3$ )	9.67 ( $\pm 1.1\text{E}-2$ )	43s
SAFE Lin (H1)	49.49	99.02	4.85	9.70	$< 10^{-4}\text{s}$
SAFE Lin (H2)	49.49	99.02	4.86	9.71	$< 10^{-4}\text{s}$
SAFE Lin (H3)	49.49	99.05	4.97	9.79	$< 10^{-4}\text{s}$
SAFE Quad 1	49.49	99.02	4.85	9.66	$< 10^{-4}\text{s}$
SAFE Quad 2	49.49	99.01	4.87	9.55	$< 10^{-4}\text{s}$
SAFE Quad 3	49.49	99.01	4.86	9.24	$< 10^{-4}\text{s}$

TABLE 5.2

Estimation of the expectations in dimension 4 with MC, MC proxy, SAFE methods (set of parameters ( $\mathcal{P}_1$ )) and execution time.

method / function	$h_1$	$h_2$	$h_3$	$h_4$	exec. time
MC	49.47 ( $\pm 1.7\text{E}-3$ )	98.45 ( $\pm 6.4\text{E}-3$ )	3.22 ( $\pm 3.2\text{E}-3$ )	18.30 ( $\pm 7.4\text{E}-3$ )	2m1s
MC Proxy	49.49 ( $\pm 1.5\text{E}-2$ )	98.50 ( $\pm 3.1\text{E}-2$ )	3.16 ( $\pm 4.0\text{E}-3$ )	18.25 ( $\pm 1.1\text{E}-2$ )	1m23s
SAFE Lin (H1)	49.48	98.48	3.17	18.28	1h16m
SAFE Lin (H2)	49.48	98.49	3.18	18.31	7s
SAFE Lin (H3)	49.48	98.50	3.23	18.47	0.3s
SAFE Quad 1	49.48	98.48	3.17	18.20	3s
SAFE Quad 2	49.48	98.48	3.17	17.98	0.2s
SAFE Quad 3	49.48	98.49	3.22	16.92	0.02s

TABLE 5.3

Estimation of the expectations in dimension 6 with MC, MC proxy and SAFE methods (set of parameters ( $\mathcal{P}_1$ )) and execution time.

method / function	$h_1$	$h_2$	$h_3$	$h_4$	exec. time
MC	49.47 ( $\pm 1.3\text{E}-3$ )	98.27 ( $\pm 5.3\text{E}-3$ )	2.47 ( $\pm 2.5\text{E}-3$ )	22.42 ( $\pm 6.6\text{E}-3$ )	2m58s
MC Proxy	49.47 ( $\pm 1.9\text{E}-2$ )	98.29 ( $\pm 3.7\text{E}-2$ )	2.43 ( $\pm 3.3\text{E}-3$ )	22.25 ( $\pm 1.3\text{E}-2$ )	2m2s
SAFE Lin (H2)	49.47	98.30	2.44	22.33	4h56m
SAFE Lin (H3)	49.48	98.31	2.49	22.54	2m7s
SAFE Quad 1	49.47	98.29	2.43	22.19	1h30m
SAFE Quad 2	49.48	98.30	2.43	21.90	1m30s
SAFE Quad 3	49.48	98.31	2.46	19.77	2s

**5.2. Results.** We study both the accuracy and the speed of the SAFE method in comparison to Monte-Carlo simulations. In Tables 5.1-5.2-5.3-5.4-5.5, we report the results in dimensions 2-4-6-8-10, with execution times, for all methods and all terminal functions using the set of parameters ( $\mathcal{P}_1$ ). In Tables 5.6-5.7 we give the results in dimension 4 using the

TABLE 5.4

Estimation of the expectations in dimension 8 with MC, MC proxy and SAFE methods (set of parameters ( $\mathcal{P}_1$ )) and execution time.

method / function	$h_1$	$h_2$	$h_3$	$h_4$	exec. time
MC	49.47 ( $\pm 1.2\text{E}-3$ )	98.17 ( $\pm 4.6\text{E}-3$ )	2.03 ( $\pm 2.1\text{E}-3$ )	25.04 ( $\pm 6.1\text{E}-3$ )	3m57s
MC Proxy	49.46 ( $\pm 2.2\text{E}-2$ )	98.18 ( $\pm 4.3\text{E}-2$ )	1.99 ( $\pm 2.9\text{E}-3$ )	24.74 ( $\pm 1.5\text{E}-2$ )	2m41s
SAFE Quad 3	49.48	98.21	2.00	21.29	3m39s

TABLE 5.5

Estimation of the expectations in dimension 10 with MC, MC proxy and SAFE methods (set of parameters ( $\mathcal{P}_1$ )) and execution time.

method / function	$h_1$	$h_2$	$h_3$	$h_4$	exec. time
MC	49.47 ( $\pm 1.0\text{E}-3$ )	98.12 ( $\pm 4.1\text{E}-3$ )	1.73 ( $\pm 1.9\text{E}-3$ )	26.93 ( $\pm 5.8\text{E}-3$ )	4m50s
MC Proxy	49.49 ( $\pm 2.4\text{E}-2$ )	98.18 ( $\pm 4.8\text{E}-2$ )	1.70 ( $\pm 2.7\text{E}-3$ )	26.52 ( $\pm 1.8\text{E}-2$ )	3m15s
SAFE Quad 3	49.47	98.15	1.69	22.35	5h49m
SAFE Quad 4	49.48	98.16	1.82	13.32	1m
SAFE Quad 5	49.48	98.17	1.60	21.05	0.39s

TABLE 5.6

Estimation of the expectations in dimension 4 with MC, MC proxy, SAFE methods (set of parameters ( $\mathcal{P}_2$ )) and execution time.

method / function	$h_1$	$h_2$	$h_3$	$h_4$	exec. time
MC	54.26 ( $\pm 1.4\text{E}-3$ )	119.15 ( $\pm 6.5\text{E}-3$ )	17.27 ( $\pm 5.3\text{E}-3$ )	34.55 ( $\pm 6.7\text{E}-3$ )	2m1s
MC Proxy	54.15 ( $\pm 2.4\text{E}-4$ )	118.56 ( $\pm 5.5\text{E}-2$ )	16.86 ( $\pm 1.3\text{E}-2$ )	32.91 ( $\pm 2.4\text{E}-2$ )	1m23s
SAFE Lin (H1)	54.20	118.70	16.91	33.01	1h16m
SAFE Lin (H2)	54.19	118.68	16.90	33.04	7s
SAFE Lin (H3)	54.18	118.68	16.91	33.21	0.3s
SAFE Quad 1	54.19	118.68	16.90	32.91	3s
SAFE Quad 2	54.19	118.66	16.89	32.63	0.2s
SAFE Quad 3	54.18	118.64	16.88	31.62	0.02s

sets of parameters ( $\mathcal{P}_2$ ) and ( $\mathcal{P}_3$ ). We finally plot in Figure 5.1 the relative errors (compared to the MC Proxy estimator) given by SAFE Lin and SAFE Quad w.r.t.  $\log(N)$ , using  $R = 5$ , in dimension 4, for the four terminal functions. All these computations have been coded in C++ on a Intel(R) Core(TM) i5 CPU@2.40GHz with 4 GB of ram.

▷ **Influence of the method.** Regarding the Tables 5.1-5.2-5.3, first we notice that the values obtained by MC and MC Proxy are close to each other (error smaller than 1%) which indicates a very good accuracy of the stochastic approximation of Theorems 2.1 and 2.3. The mesh size  $\delta$  is larger for SAFE Quad than for SAFE Lin; nevertheless we remark that SAFE Lin and Quad give close results and give good deterministic approximations of the values obtained by MC Proxy (the value towards which we expect convergence) and MC (the target value). Their relative errors in comparison to MC Proxy are often smaller than 1%.

TABLE 5.7

Estimation of the expectations in dimension 4 with MC, MC proxy, SAFE methods (set of parameters ( $\mathcal{P}_3$ )) and execution time.

method / function	$h_1$	$h_2$	$h_3$	$h_4$	exec. time
MC	73.08 ( $\pm 8.1\text{E}-4$ )	272.43 ( $\pm 1.1\text{E}-2$ )	100.00 ( $\pm 4.1\text{E}-3$ )	113.76 ( $\pm 5.8\text{E}-3$ )	2m1s
MC Proxy	73.09 ( $\pm 8.3\text{E}-4$ )	272.44 ( $\pm 1.1\text{E}-2$ )	100.00 ( $\pm 4.2\text{E}-3$ )	113.75 ( $\pm 6.0\text{E}-3$ )	1m23s
SAFE Lin (H1)	73.07	272.37	99.97	113.71	7h10m
SAFE Lin (H2)	73.07	272.38	99.98	113.73	35s
SAFE Lin (H3)	73.08	272.41	99.99	113.81	0.6s
SAFE Quad 1	73.08	272.41	99.99	113.67	3s
SAFE Quad 2	73.08	272.40	99.99	113.54	0.2s
SAFE Quad 3	73.07	272.39	99.98	113.04	0.02s

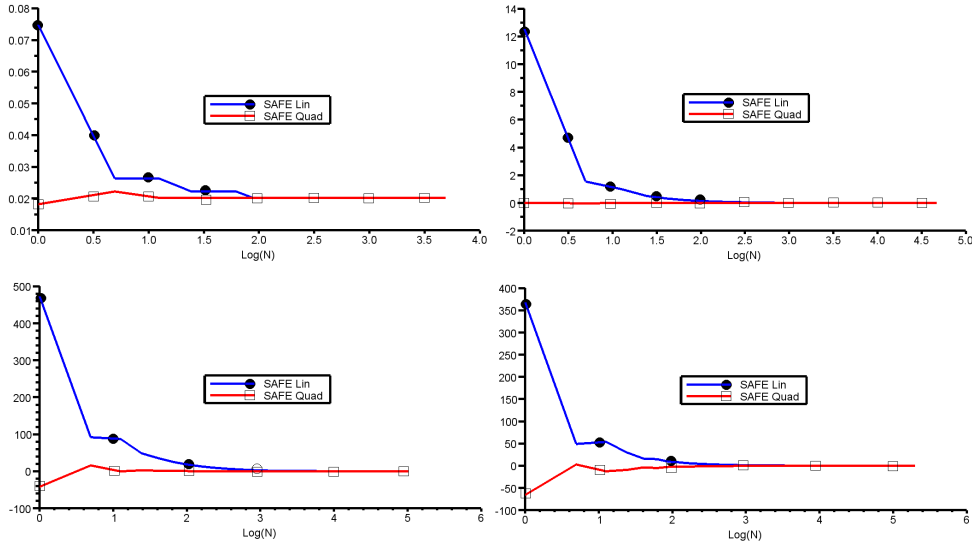


FIG. 5.1. Relative errors (in %) in comparison to MC Proxy of SAFE Lin and Quad w.r.t.  $\log(N)$  in dimension 4 (from left to right and from top to bottom: terminal functions  $h_1$ ,  $h_2$ ,  $h_3$  and  $h_4$ ).

▷ **Influence of the number of points.** We observe in Tables 5.1-5.2-5.3 that as expected, with fewer points (SAFE Lin (H3) or Quad 3), we globally lose accuracy.

▷ **Influence of the dimension and the terminal function.** Generally speaking, for  $h_1, h_2, h_3$  the accuracy is very good whatever the dimension is. For  $h_4$ , results get worse as  $d$  increases. We notice in Tables 5.3-5.4-5.5 that in high dimension, the SAFE methods give more accurate results for  $h_3$  (probably because the average of the r.v. induces smaller fluctuations) but less accurate for  $h_4$  (certainly due to the regularity breakdown of the function max in high-dimension).

▷ **Speed results.** Regarding the execution times, we notice that in dimension 2, the use of SAFE is almost instantaneous versus 1 or 2 minutes for accurate Monte-Carlo simulations. In dimension 4, it takes 1h16 for SAFE Lin (H1), 7s for SAFE Lin (H2) and less for the other SAFE methods. Even if 1h16 seems to be quite important, it is still faster than MC (taking

into account the discretization effort) but slower than MC Proxy. In dimension 6, SAFE Lin (H2) takes almost 5h which is comparable to MC. SAFE Quad 1 takes 1h30 but SAFE Lin (H3) (2m7s), Quad 2 and 3 (1m30s and 2s) are very competitive. Notice that in dimension 5 (which results are not shown here), SAFE Lin (H2) and Quad 1 need approximatively 4m30s and 2m, a performance which is close to MC Proxy. In dimension 7 (not presented here), SAFE Lin (H3), Quad 2 and 3 respectively necessitate 47m, 34m and 21s. In higher dimension ( $\geq 8$ ), only SAFE Quad 3 is competitive with very satisfying results except for the max terminal function. It takes 3m39s in dimension 8 (close to MC Proxy), about 36m in dimension 9 (not reported here) and almost 6h in dimension 10 (comparable to MC). Of course, this huge amount of computations can be easily split among parallel processors: this is a nice feature of the SAFE method. Also, the efficiency may be much improved by using sparse grids [5], this is left to further investigation. With very few points, SAFE Quad 4–5 give in dimension 10 very good results for  $h_1$  and  $h_2$ , good results for  $h_3$  (relative errors of order 5%) and pretty poor results with  $h_4$ , for execution times of respectively 1m and less than 1s. This shows that for smooth functions, we may reach very high accuracies with few points.

▷ **Convergence results.** We notice from Figure 5.1 that the convergence is quite fast for both SAFE Lin and Quad. Moreover the more regular the terminal function, the faster the convergence. For large regularity of  $h$ , only few points are needed to achieve convergence. On the contrary for the function max, the convergence speeds are similar (taking into account that the computational cost at fixed  $N$  is higher for SAFE Quad).

▷ **Influence of the drift and the initial point.** Table 5.6 (parameters ( $\mathcal{P}_2$ )) shows that the accuracy is worsened as the drift gets larger, inaccuracies increasing with the irregularity of the terminal function. This transport term probably increases the deviations of the diffusion and worsens the accuracy of the proxy. On the contrary, considering an initial point at 1 which induces for  $\sigma$  small variations and magnitude leads to better results as presented in Table 5.7 (parameters ( $\mathcal{P}_3$ )), notably for the function  $h_3$ . All these observations are coherent with our error estimates in Theorems 2.1–2.6–2.8.

**Appendix A. Computation of the correction terms in Theorems 2.1 and representation as sensitivities.** We follow the routine of [9], with some adaptations due the multidimensional setting. We first provide integration by parts formulas useful for the explicit derivation of the correction terms. In the following,  $c_t : [0, T] \rightarrow \mathbb{R}^{1 \times q}$ ,  $C_t : [0, T] \rightarrow \mathbb{R}^{d \times q}$  are square integrable and predictable processes,  $A_t : [0, T] \rightarrow \mathbb{R}^{d \times q}$  is a square integrable and deterministic process and  $\psi, \psi_1, \dots, \psi_d : \mathbb{R}^d \rightarrow \mathbb{R}$  are smooth functions with bounded derivatives.

LEMMA A.1. *One has:*

$$(A.1) \quad \mathbb{E}[\psi(\int_0^T A_t dW_t) \int_0^T c_t dW_t] = \mathbb{E}[\nabla \psi(\int_0^T A_t dW_t) \int_0^T A_t c_t^* dt],$$

$$(A.2) \quad \mathbb{E}[(\psi_1(\int_0^T A_t dW_t), \dots, \psi_d(\int_0^T A_t dW_t)) \int_0^T C_t dW_t] = \sum_{i,j=1}^d \mathbb{E}[\partial_{x_j} \psi_i(\int_0^T A_t dW_t) \int_0^T (A_t C_t^*)_i^j dt].$$

*Proof.* We begin with (A.1). The process  $A$  being deterministic, the Malliavin derivative  $D_s \int_0^T A_t dW_t$  is equal to  $A_s \mathbb{1}_{s \leq T}$  and  $\psi(\int_0^T A_t dW_t) \in \mathbb{D}^{1,\infty}$  with  $D_s[\psi(\int_0^T A_t dW_t)] = \psi'(\int_0^T A_t dW_t) A_s \mathbb{1}_{s \leq T}$ . Then identify  $\int_0^T c_t dW_t$  with the Skorohod operator and apply the duality relationship [20, Definition 1.3.1 and Proposition 1.3.11]. To derive (A.2), apply (A.1)

with  $c_t = C_t^i$  for any  $i \in \{1, \dots, d\}$ :

$$\begin{aligned} \mathbb{E}[(\psi_1(\int_0^T A_t dW_t), \dots, \psi_d(\int_0^T A_t dW_t)) \int_0^T C_t dW_t] &= \sum_{i=1}^d \mathbb{E}[\nabla \psi_i(\int_0^T A_t dW_t) \int_0^T A_t (C_t^i)^* dt] \\ &= \sum_{i,j=1}^d \mathbb{E}[\partial_{x_j} \psi_i(\int_0^T A_t dW_t) \int_0^T (A_t C_t^*)_{ij}^j dt]. \end{aligned}$$

□

We are now in a position to prove:

**PROPOSITION A.2.** *Assume  $(\mathcal{H}_{x_0})$ . For any smooth function  $\phi : \mathbb{R}^d \mapsto \mathbb{R}$  with bounded derivatives, we have:*

$$\mathbb{E}[\nabla \phi(X_T^P) \dot{X}_T] = \text{Cor}_{2,\phi},$$

where  $\text{Cor}_{2,\phi}$  is defined in (2.7).

*Proof.* In the proof, we repeatedly use the identity  $\mathbb{E}(\partial^\alpha \phi(X_T^P)) = \partial^\alpha \bar{\phi}^P(0)$  for any multi-index  $\alpha$ . In view of (2.5) and (2.4), one has to transform  $\mathbb{E}[\nabla \phi(X_T^P) \dot{X}_T] = I_1 + I_2$  with

$$I_1 := \mathbb{E}[\nabla \phi(X_T^P) \int_0^T b'_t(X_t^P - x_0) dt], \quad I_2 := \mathbb{E}[\nabla \phi(X_T^P) \int_0^T \sum_{j=1}^d \sigma'_{j,t}(X_t^P - x_0) dW_t^j].$$

We begin with  $I_1$ . Writing  $X_t^P - x_0 = \int_0^t \sigma_s dW_s + \int_0^t b_s ds$ , we obtain  $I_1 = I_{1a} + I_{1b}$  with

$$\begin{aligned} I_{1a} &:= \mathbb{E}[\nabla \phi(X_T^P) \int_0^T b'_t(\int_0^t b_s ds) dt] = \nabla \bar{\phi}^P(0) \int_0^T b'_t(\int_0^t b_s ds) dt, \\ I_{1b} &:= \mathbb{E}[\nabla \phi(X_T^P) \int_0^T b'_t(\int_0^t \sigma_s dW_s) dt] = \int_0^T \mathbb{E}[\nabla \phi(X_T^P) (\int_0^t b'_t \sigma_s dW_s)] dt. \end{aligned}$$

The contribution  $I_{1a}$  is in its final form. Regarding  $I_{1b}$ , for any  $t \in [0, T]$  apply formula (A.2) with  $(\psi_1, \dots, \psi_d)(x) = \nabla \phi(m_T^P + x)$ ,  $(A_s)_{s \in [0, T]} = (\sigma_s)_{s \in [0, T]}$  and  $(C_s)_{s \in [0, T]} = (\mathbb{1}_{s \leq t} b'_t \sigma_s)_{s \in [0, T]}$  to directly obtain:

$$I_{1b} = \mathbb{E}[\sum_{i,j=1}^d \partial_{x_i, x_j}^2 \phi(X_T^P) \int_0^T (\int_0^t \sigma_s \sigma_s^* (b'_t)^* ds)_i^j dt] = \sum_{i,j=1}^d \partial_{\epsilon_i, \epsilon_j}^2 \bar{\phi}^P(0) \int_0^T (b'_t)^i (\int_0^t \Sigma_{j,s} ds) dt.$$

We now handle the term  $I_2$ . Let  $D_t$  be the  $d$ -dimensional square matrix whose  $j$ -th column is equal to  $\sigma'_{j,t}(X_t^P - x_0)$ : using again (A.2), we derive

$$\begin{aligned} I_2 &= \mathbb{E}[\nabla \phi(X_T^P) \int_0^T D_t dW_t] = \mathbb{E}[\sum_{i,j=1}^d \partial_{x_i, x_j}^2 \phi(X_T^P) \int_0^T (\sigma_t D_t^*)_{ij}^j dt] \\ &= \mathbb{E}[\sum_{i,j,k=1}^d \partial_{x_i, x_j}^2 \phi(X_T^P) \int_0^T \sigma_{k,t}^j D_{k,t}^i dt] = \mathbb{E}[\sum_{i,j,k,l=1}^d \partial_{x_i, x_j}^2 \phi(X_T^P) \int_0^T [\sigma_k^j \partial_{x_l} \sigma_k^i](t, x_0) (X_t^P - x_0)^l dt]. \end{aligned}$$

Thanks to the symmetry of the Hessian matrix  $H(\phi)$ , we also have

$$\begin{aligned} I_2 &= \frac{1}{2} \mathbb{E}[\sum_{i,j,k,l=1}^d \partial_{x_i, x_j}^2 \phi(X_T^P) \int_0^T [\sigma_k^i \partial_{x_l} \sigma_k^j + \sigma_k^j \partial_{x_l} \sigma_k^i](t, x_0) (X_t^P - x_0)^l dt] \\ &= \frac{1}{2} \mathbb{E}[\sum_{i,j,l=1}^d \partial_{x_i, x_j}^2 \phi(X_T^P) \int_0^T \partial_{x_l} \Sigma_{j,t}^i(t, x_0) (X_t^P - x_0)^l dt] = \frac{1}{2} \mathbb{E}[\sum_{i,j=1}^d \partial_{x_i, x_j}^2 \phi(X_T^P) \int_0^T (\Sigma_{j,t}^i)'(X_t^P - x_0) dt]. \end{aligned}$$



From this point, the computations are similar to those for  $I_1$ ; briefly, writing  $X_t^P - x_0 = \int_0^t \sigma_s dW_s + \int_0^t b_s ds$ , we decompose  $I_2 = \frac{1}{2}(I_{2a} + I_{2b})$  with

$$I_{2a} := \mathbb{E} \left[ \sum_{i,j=1}^d \partial_{x_i, x_j}^2 \phi(X_T^P) \int_0^T (\Sigma_{j,t}^i)' \left( \int_0^t b_s ds \right) dt \right] = \sum_{i,j=1}^d \partial_{\epsilon_i, \epsilon_j}^2 \bar{\phi}^P(0) \int_0^T (\Sigma_{j,t}^i)' \left( \int_0^t b_s ds \right) dt,$$

$$I_{2b} := \mathbb{E} \left[ \sum_{i,j=1}^d \partial_{x_i, x_j}^2 \phi(X_T^P) \int_0^T (\Sigma_{j,t}^i)' \left( \int_0^t \sigma_s dW_s \right) dt \right] = \int_0^T \mathbb{E} \left[ \sum_{i,j=1}^d \partial_{x_i, x_j}^2 \phi(X_T^P) \left( \int_0^t (\Sigma_{j,t}^i)' \sigma_s dW_s \right) \right] dt.$$

Then use (A.1) and the symmetry of  $\Sigma_t$  in order to get

$$\begin{aligned} I_{2b} &= \int_0^T \mathbb{E} \left[ \sum_{i,j=1}^d \nabla(\partial_{x_i, x_j}^2 \phi)(X_T^P) \left( \int_0^t \sigma_s [(\Sigma_{j,t}^i)' \sigma_s]^* ds \right) \right] dt \\ &= \mathbb{E} \left[ \sum_{i,j=1}^d \nabla(\partial_{x_i, x_j}^2 \phi)(X_T^P) \int_0^T \left( \int_0^t \Sigma_s ds \right) ((\Sigma_{j,t}^i)')^* dt \right] \\ &= \sum_{i,j,k=1}^d \partial_{\epsilon_i, \epsilon_j, \epsilon_k}^3 \bar{\phi}^P(0) \left[ \int_0^T \left( \int_0^t \Sigma_s ds \right) ((\Sigma_{j,t}^i)')^* dt \right]^k = \sum_{i,j,k=1}^d \partial_{\epsilon_i, \epsilon_j, \epsilon_k}^3 \bar{\phi}^P(0) \int_0^T (\Sigma_{j,t}^i)' \left( \int_0^t \Sigma_{k,s} ds \right) dt. \end{aligned}$$

Gathering all the contributions achieves the proof.  $\square$

#### REFERENCES

- [1] V. BALLY AND G. PAGÈS, *A quantization algorithm for solving multi-dimensional optimal stopping problems*, Bernoulli, 9 (2003), pp. 1003–1049.
- [2] V. BALLY AND D. TALAY, *The law of the Euler scheme for stochastic differential equations: I. Convergence rate of the distribution function*, Probab. Theory Related Fields, 104-1 (1996), pp. 43–60.
- [3] R. BOMPIS AND E. GOBET, *Asymptotic and non asymptotic approximations for option valuation*, in Recent Developments in Computational Finance: Foundations, Algorithms and Applications, T. Gerstner and P. Kloeden, eds., World Scientific Publishing Company, 2012.
- [4] S. BRENNER AND L. SCOTT, *The mathematical theory of finite element methods*, Texts in Applied Mathematics 15. New York, NY: Springer, third ed., 2008.
- [5] H.-J. BUNGARTZ AND M. GRIEBEL, *Sparse grids*, Acta Numerica, 13 (2004), pp. 147–269.
- [6] D. DUFFIE AND P. GLYNN, *Efficient Monte Carlo simulation of security prices*, Ann. Appl. Probab., 5 (1995), pp. 897–905.
- [7] J. FOUQUE, G. PAPANICOLAOU, R. SIRCAR, AND S. KNUT, *Multiscale stochastic volatility for equity, interest rate, and credit derivatives*, Cambridge University Press, Cambridge, 2011.
- [8] E. GOBET, *Euler schemes and half-space approximation for the simulation of diffusions in a domain*, ESAIM: Probability and Statistics, 5 (2001), pp. 261–297.
- [9] E. GOBET AND M. MIRI, *Weak approximation of averaged diffusion processes*, To appear in Stochastic Processes and their Applications, (2013).
- [10] S. GRAF AND H. LUSCHGY, *Foundations of quantization for probability distributions*, vol. 1730 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 2000.
- [11] I. KARATZAS AND S. SHREVE, *Brownian motion and stochastic calculus*, Springer-Verlag, 1991.
- [12] A. KEBAIER, *Statistical romberg extrapolation: a new variance reduction method and applications to option pricing*, The Annals of Applied Probability, 15 (2005), pp. 2681–2705.
- [13] P. KLOEDEN AND E. PLATEN, *Numerical solution of stochastic differential equations. 4th corrected printing*, Applications of Mathematics 23. Berlin: Springer, 2010.
- [14] H. KUNITA, *Stochastic flows and stochastic differential equations*, Cambridge Studies in Advanced Mathematics. 24. Cambridge: Cambridge University Press, 1997.
- [15] S. KUSUOKA, *Approximation of expectation of diffusion processes based on Lie algebra and Malliavin calculus*. Kusuoka, Shigeo (ed.) et al., Advances in mathematical economics. Vol. 6. Tokyo: Springer. Adv. Math. Econ. 6, 69-83 (2004), 2004.
- [16] J. LEMOR, E. GOBET, AND X. WARIN, *Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations*, Bernoulli, 12 (2006), pp. 889–916.

- [17] M. LORIG, S. PAGLIARANI, AND A. PASCUCCI, *Implied vol for any local-stochastic vol model*, Preprint SSRN, (2013).
- [18] T. LYONS AND N. VICTOIR, *Cubature on Wiener space*, Proc. R. Soc. Lond., Ser. A, Math. Phys. Eng. Sci., 460 (2004), pp. 169–198.
- [19] S. NINOMIYA AND N. VICTOIR, *Weak approximation of stochastic differential equations and application to derivative pricing*, Appl. Math. Finance, 15 (2008), pp. 107–121.
- [20] D. NUALART, *Malliavin calculus and related topics*, Springer-Verlag, Berlin, second ed., 2006.
- [21] D. TALAY AND L. TUBARO, *Expansion of the global error for numerical schemes solving stochastic differential equations*, Stochastic Analysis and Applications, 8-4 (1990), pp. 94–120.